2017

Deciphering Chronometabolic Dynamics Through Metabolomics, Stable Isotope Tracers, And Genome-Scale Reaction Modeling

Seth D. Rhoades
University of Pennsylvania, srhoades10@gmail.com

Deciphering Chronometabolic Dynamics Through Metabolomics, Stable Isotope Tracers, And Genome-Scale Reaction Modeling

Abstract
Synchrony across environmental cues, endogenous genetic clocks, sleep/wake cycles, and metabolism evoke physiological harmony for organismal health. Perturbation of this synchrony has been recently correlated with a growing list of pathologies, which is alarming given the ubiquity of sleep deprivation, mistimed light exposure, and altered eating schedules in modern society. Deeper insights into clocks, sleep, and metabolism are necessary to understand these outcomes. In this work, extensive metabolic profiles of circadian systems were obtained from the development of new liquid chromatography mass spectrometry (LC-MS) metabolomics methods. These methods were applied to Drosophila melanogaster to discern relative influences of environmental and genetic drivers of metabolic cycles. Unique sets of metabolites oscillated with 24-hour circadian periods under light:dark (LD) and constant darkness (DD) conditions, and ultradian rhythms were noted for clock mutant flies under LD, suggesting clock-independent metabolic cycles driven by environmental inputs. However, this metabolomic analysis does not fully capture the inherently dynamic nature of circadian metabolism. These LC-MS methods were adapted to analyze isotope enrichments from a novel 13C6 glucose injection platform in Drosophila. Metabolic flux cycles were noted from glucose carbons into serine, glutamine and reduced glutathione biosynthesis, and altered under sleep deprivation, demonstrating unique energy and redox demands in perturbed sleep/wake cycles. Global isotopolome shifts were most notable in WT flies after lights-on, suggesting a catabolic rush from glucose oxidation early in the active phase. As the scope of these isotope tracer-based metabolomic analyses expand, attributing labeling patterns to specific reactions requires consideration of genome-scale metabolic networks. A new computational approach was developed, called the IsoPathFinder, which uncovered biosynthetic paths from glucose to serine, and extends to glycine and glutathione production. Carbon flux into glutamine was predicted to occur through the TCA cycle, supported by enzyme thermodynamics and circadian expression datasets. This tool is presented as a new mechanism to simulate additional isotope tracer experiments, with broad applicability beyond circadian research. Collectively, a new set of analytical and computational tools are developed to both produce dynamic metabolomic data and improve data interpretability, with applications to uncover new chronometabolic connections.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Pharmacology

First Advisor
Aalim M. Weljie

DECIPHERING CHRONOMETABOLIC DYNAMICS THROUGH METABOLOMICS, STABLE ISOTOPE TRACERS, AND GENOME-SCALE REACTION MODELING

Seth D. Rhoades

A DISSERTATION

in

Pharmacology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Aalim M. Weljie, Ph.D., Research Assistant Professor of Pharmacology

Graduate Group Chairperson

_____

Julie Blendy, Ph.D., Professor of Pharmacology

Dissertation Committee

Ian Blair, Ph.D., A.N. Richards Professor of Pharmacology

Harry Ischiropoulos, Ph.D., Research Professor of Pediatrics

Zachary Schug, Ph.D., Assistant Professor, Molecular and Cellular Oncogenesis

Amita Sehgal, Ph.D., John Herr Musser Professor of Neuroscience

DECIPHERING CHRONOMETABOLIC DYNAMICS THROUGH METABOLOMICS, STABLE ISOTOPE TRACERS, AND GENOME-SCALE REACTION MODELING

DEDICATION

This work is dedicated to my parents, Donna and Bruce. Words cannot adequately describe the love you have demonstrated in working endlessly to provide me opportunities to actualize my ambitions, and the deep thankfulness I feel for teaching me about grit and grace, ethic and ethics, and passion and compassion.

“The greatest obstacle to discovery is not ignorance – it is the illusion of knowledge”

-Daniel J. Boorstin

ACKNOWLEDGMENT

This work, and my training at the University of Pennsylvania, is the product of an unparalleled web of collaboration. I have garnered both technical expertise and broadened perspectives on performing impactful research, means of experimentation, and the future of therapeutic discovery from excellent scientists too numerous to fully mention here. I would especially like to thank Dr. Aalim Weljie, who took me in as his first doctoral student and demonstrated extreme patience with me as I attempted to learn about new fields of metabolomics and circadian rhythms. I am thankful for the space I was given to explore new techniques and refine my own skills and interests, even if not always aligned with the lab objectives. I also want to thank members of the Weljie lab, including Emily Mackay, Barry Slaff, Sidney Bolden, Saikumari Krishnaiah, and Arjun Sengupta, for their help and support.

Of the numerous collaborations I maintained in the Weljie lab, the two most notable scientists who contributed substantially to my training are Drs. Amita Sehgal and Garret FitzGerald. I benefited greatly from the shared resources through Amita's lab, with particular help from Paula Haynes and Shirley Zhang, who aided many technical aspects of my experiments, and I received tremendous guidance out of the regular with Amita, who possesses prodigious knowledge of clocks and sleep. I am also extremely fortunate to work alongside Garret and members of his lab. I have never met someone with such perspective on the impact of our work as scientists, and foresight to the future of medicine. I am additionally grateful to participate in a collaboration with Dr. Carsten Skarke, comprising work which sparked my career interests.

I also want to thank my thesis committee, Drs. Ian Blair, Harry Ischiropoulos, Amita Sehgal, and Zachary Schug, who provided helpful guidance in my project. Finally, I am thankful for the support and sincere friendships I have fostered with classmates in the Pharmacology Graduate Group, as I would not have made it through graduate school without them.

ABSTRACT

DECIPHERING CHRONOMETABOLIC DYNAMICS THROUGH METABOLOMICS, STABLE ISOTOPE TRACERS, AND GENOME-SCALE REACTION MODELING

Seth D. Rhoades

Aalim M. Weljie, Ph.D.

Synchrony across environmental cues, endogenous genetic clocks, sleep/wake cycles, and metabolism evoke physiological harmony for organismal health. Perturbation of this synchrony has been recently correlated with a growing list of pathologies, which is alarming given the ubiquity of sleep deprivation, mistimed light exposure, and altered eating schedules in modern society. Deeper insights into clocks, sleep, and metabolism are necessary to understand these outcomes. In this work, extensive metabolic profiles of circadian systems were obtained from the development of new liquid chromatography mass spectrometry (LC-MS) metabolomics methods. These methods were applied to *Drosophila melanogaster* to discern relative influences of environmental and genetic drivers of metabolic cycles. Unique sets of metabolites oscillated with 24-hour circadian periods under light:dark (LD) and constant darkness (DD) conditions, and ultradian rhythms were noted for clock mutant flies under LD, suggesting clock-independent metabolic cycles driven by environmental inputs. However, this metabolomic analysis does not fully capture the inherently dynamic nature of circadian metabolism. These LC-MS methods were adapted to analyze isotope enrichments from a novel $^{13}C_6$ glucose injection platform in *Drosophila*. Metabolic flux cycles were noted from glucose carbons into serine, glutamine and reduced glutathione biosynthesis, and altered under sleep deprivation, demonstrating unique energy and redox demands in perturbed sleep/wake cycles. Global isotopolome shifts were most notable in WT flies after lights-on, suggesting a catabolic rush from glucose oxidation early in the active phase. As the scope of these isotope tracer-based metabolomic analyses expand, attributing labeling patterns to specific reactions requires consideration of genome-scale metabolic networks. A new computational approach was developed, called the IsoPathFinder,

which uncovered biosynthetic paths from glucose to serine, and extends to glycine and glutathione production. Carbon flux into glutamine was predicted to occur through the TCA cycle, supported by enzyme thermodynamics and circadian expression datasets. This tool is presented as a new mechanism to simulate additional isotope tracer experiments, with broad applicability beyond circadian research. Collectively, a new set of analytical and computational tools are developed to both produce dynamic metabolomic data and improve data interpretability, with applications to uncover new chronometabolic connections.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2 – Comprehensive optimization of LC-MS metabolomics methods using design of experiments (COLMeD)

CHAPTER 3 – Circadian- and light-driven metabolic rhythms in *Drosophila melanogaster*

CHAPTER 4 – *In vivo* circadian isotopolomics platform uncovers unique diurnal flux patterns perturbed under sleep deprivation

CHAPTER 5 – *In silico* isotopologue-mapping to trace metabolic pathways at genome-scale

LIST OF ILLUSTRATIONS

CHAPTER 1 – Introduction

CHAPTER 2 – Comprehensive optimization of LC-MS metabolomics methods using design of experiments (COLMeD)

CHAPTER 3 – Circadian- and light-driven metabolic rhythms in *Drosophila melanogaster*

CHAPTER 4 – *In vivo* circadian isotopolomics platform uncovers unique diurnal flux patterns perturbed under sleep deprivation

CHAPTER 5 – *In silico* isotopologue-mapping to trace metabolic pathways at genome-scale

LIST OF ABBREVIATIONS

AKH, Adipokinetic hormone

AMP, Adenosine monophosphate

AMPK, AMP-dependent protein kinase

ANOVA, Analysis of variance

ATP, Adenosine triphosphate

AUC, Area under the curve

AVP, Arginine vasopressin

CAS, Chemical abstracts service

CBM, Constraint-based modeling

CCF, Composite face-centered

CCG, Clock-controlled gene

CE, Capillary electrophoresis

CK2, Casein kinase 2

CLK, Clock

COLMeD, Comprehensive optimization of LC-MS metabolomics methods using design of experiments

CRH, Corticotrophin-releasing hormone

CRY, Cryptochrome

CV, Cross-validation

CYC, Cycle

DAG, Diacylglycerol

DBT, Doubletime

DD, Dark:dark

dILP, Drosophila insulin-like-peptide

DN, dorsal neuron

DoE, Design of experiments

EC, Enzyme commission

ESI, Electrospray ionization

GABA, Gamma-aminobutyric acid

GC, Gas chromatography

GlcN6P, Glucosamine-6-phosphate

GlcNAc, N-Acetylglucosamine

GSH, Reduced glutathione

GSK3B, Glycogen synthase kinase 3 beta

HFD, High fat diet

HILIC, Hydrophilic interaction liquid chromatography

HPA, Hypothalamic-pituitary-adrenal

LC, Liquid chromatography

LD, Light:dark

LN, Lateral neuron

LOESS, Locally-weighted scatterplot smoothing

MDV, Mass distribution vector

m/z, Mass-to-charge

MLR, Multiple linear regression

MRM, Multiple reaction monitoring

MS, Mass spectrometry

MS/MS, Tandem mass spectrometry

NAD+, Nicotinamide adenine dinucleotide

NAMPT, Nicotinamide phosphoribosyl transferase

NMR, Nuclear magnetic resonance

OPLS-DA, Orthogonal partial least squares discriminant analysis

PBS, Phosphate-buffered saline

PCA, Principal components analysis

PER, Period

PPAR, Peroxisome proliferator-activated receptor

PPP, Pentose phosphate pathway

PTM, Post-translational modification

PVN, Paraventricular nucleus

QC, Quality control

QqQ, Triple quadrupole

qTOF, Quadrupole time-of-flight

ROR, receptor-related orphan receptor

ROS, Reactive oxygen species

RPLC, Reverse-phase liquid chromatography

RSD, Relative standard deviation

SCN, Suprachiasmatic nucleus

SIRT, Sirtuin

SLIMB, Supernumerary limbs

SMILES, Simplified molecular-input line-entry system

S/N, Signal-to-noise

TAG, Triglyceride

TCA, Tricarboxylic acid

TIM, Timeless

TLC, Thin-layer chromatography

TOF, Time-of-flight

tRF, Time-restricted feeding

UPLC, Ultra-performance liquid chromatography

VIP, Variable importance

VIP, Vasoactive intestinal peptide

ZT, Zeitgeber time

CHAPTER 1 - Introduction

1.1 Intersection of circadian rhythms, sleep, and metabolism in health and disease

History and characterization of circadian rhythms

The ancient Greek philosopher Heraclitus famously coined the phrase '*panta rhei*', or 'everything flows', which has since been expounded into the aphorism 'the only thing constant is change'. Indeed, life is flux for all organisms on this planet, encountering daily changes in light, temperature, and food availability. Amidst a complex external environment, emergent patterns arise from which most organisms have adapted mechanisms to internally anticipate and synchronize with predictable environmental changes, ultimately conferring a survival advantage. At the heart of these internal timekeeping mechanisms lie the circadian clock. This intrinsic biological clock was deduced experimentally after the observations of daily leaf movements in heliotrope plants under constant darkness by Jean-Jacques d'Ortous de Mairan in 1729 (de Marian, 1729), leading to the development of the basic tenants of biological clocks by Pittendrigh and Aschoff, namely persistent cyclic periods under constant conditions, entrainment to environmental signals, and robustness against temperature changes (Pittendrigh, 1960, Aschoff, 1965). Strictly circadian processes refer to those which take place over a roughly 24-hour period (Latin "circa diem" – about a day), although shorter periods are also known to exist (termed ultradian rhythms). In higher organisms, molecular clocks consist of a hierarchy structure, where input pathways receive and transmit environmental cues to a central oscillator, which in turn generates rhythmic outputs to govern many metabolic, physiological, and behavioral processes at an organismal level. For instance, mammals harbor a set of light-responsive retinal ganglion cells, which project to the central pacemaker, the suprachiasmatic nucleus (SCN), via the retinohypothalamic tract (Moore, 1982). Upon activation, the SCN expresses arginine vasopressin (AVP) and vasoactive intestinal peptide (VIP), which are thought to be the main synchronizing agents which correspond to a robust cyclic output, with projections to numerous nuclei in the hypothalamus and thalamus (Maywood et al., 2011). These rhythmic outputs can

synchronize circadian oscillators in other parts of the organism, commonly called the peripheral clocks. The SCN can thus act as the conductor to the orchestra of cells and organs across both the central nervous system and the periphery, evoking physiologic harmony during times of organismal synchrony.

Our current understanding in the molecular basis of clocks originated from mutagenesis studies in *Drosophila melanogaster* (Konopka and Benzer, 1971), fueling the discovery of autoregulatory transcription-translation feedback loops (Glossop and Hardin, 2002). The functional orthologs to *Drosophila* clocks have since been discovered in mammals, and have been reviewed extensively (Lowrey and Takahashi, 2011, Figure 1.1). At the core of the mammalian clock, proteins CLOCK and BMAL1 dimerize and bind to E-box promoter elements to activate numerous clock-controlled genes (CCGs). Repressor proteins such as cryptochrome and period (CRY1-2 and PER1-3) translocate back into the nucleus to suppress the CLOCK:BMAL1 complex, thus completing the feedback loop. Parallel loops also exist to control Bmal1 transcription. In particular, BMAL1 drives expression of Rev-Erbα/β and retinoic acid receptor-related orphan receptors (RORs) which compete for binding to ROR elements to repress or activate Bmal1 expression, respectively. Beyond these core loops, there exists a remarkable complexity of additional inputs to the clock, including a plethora of post-translational modifications (PTMs), protein degradation pathways, nuclear hormone receptors, metabolism, and epigenetics, which can all couple to our most basic phenotypes of sleep and wake cycles.

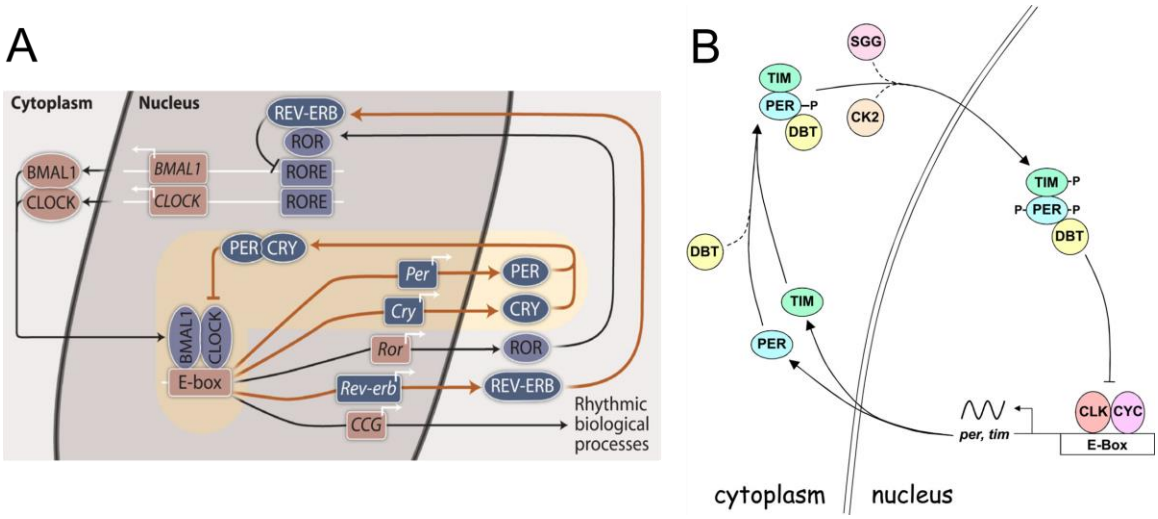


Figure 1.1 Transcription-translation feedback loops of the core mammalian (A) and *Drosophila* (B) circadian clocks. Mammalian circadian clock reprinted with permission from Yang et al., 2013; *Drosophila* circadian clock reprinted with permission from Hardin, 2002.

Circadian clocks and sleep

Sleep, broadly defined as a period of inactivity with an increased arousal threshold, stereotypical body positions, and sleep rebound following periods of disruption or deprivation (Hendricks et al., 2000), is a highly conserved process with two major regulatory mechanisms. One mechanism, the S process, is a homeostatic drive for sleep which increases throughout the active phase (i.e. the light phase for diurnal organisms) and is alleviated upon sleep onset (Borbely, 1982). This process is empirically observed and understood, albeit with little mechanistic underpinning. The other major mechanism, the C process, is the circadian control over sleep/wake behaviors. Melatonin exhibits a robust diurnal secretion from the pineal gland, driven by SCN and paraventricular nucleus (PVN) projections to elicit sleep (Teclemariam-Mesbah et al., 1999), which is countered by cortisol rhythms, from pulsatile secretions of corticotropin-releasing hormone (CRH) out of the PVN to elicit morning arousal (Spencer et al., 1998). Perturbations in these cycles, whether from suppressed melatonin production by light exposure at night, or a perturbed hypothalamic-pituitary-adrenal (HPA) axis from stress or pathologies, result in fragmented or lost sleep (Balbo et al., 2010).

Mutagenesis studies have yielded few genes which exert a dominant control on sleep quality and quantity, implying a complex genetic sleep architecture (Yamamoto et al., 2008). However, some notable polymorphisms in core clock genes impact sleep, including associations of Per variants with sleep timing (Carpen et al., 2006, Archer et al., 2003). Additionally, genetic mouse models of Clock, Bmal1, and Cry1/2 display altered sleep duration (Naylor et al., 2000, Wisor et al., 2002, Laposky et al., 2005). These tantalizing connections of sleep phenotypes and circadian clocks warrant deeper mechanistic experimentation, wherein lies an inextricable connection of clocks with hormone production and metabolism.

Circadian clocks and metabolism

Circadian clocks and metabolism exhibit a complex bidirectional control, and has recently been a subject of intense research in the circadian field. Transcriptomics in mice have

demonstrated a striking enrichment of rhythms in rate-limiting enzymes across major metabolic pathways, including glucose, fatty acid, amino acid, and bile acid metabolism (Panda et al., 2002). Intriguingly, expression of many of these metabolic enzymes can display ultradian rhythms of 8 or 12 hours (Hughes et al., 2009), which may align with shorter temporal patterns of feeding and behavior. In addition to these extrinsic factors, there is also evidence that the cell-autonomous circadian clocks drive metabolic oscillations. For example, corticosterone rhythms were abolished in SCN-lesioned rats (Moore and Eichler, 1972), and circulating glucose, triglycerides, and fatty acids have been shown to oscillate under fasting or fixed conditions such as forced wakefulness or constant light (Scheer et al., 2009, la Fleur et al., 1999 and Dallmann et al., 2012). Additionally, nicotinamide phosphoribosyl-transferase (NAMPT), the rate-limiting step in the nicotinamide adenine dinucleotide (NAD+) salvage pathway, exhibits a strong circadian rhythm, particularly in the liver (Ramsey et al., 2009). This clock-metabolism connection has several critical effects, including an intracellular oscillation in redox potential, noted in NAD+/NADH cycles. Redox rhythms can also exist as part of a highly conserved and ancient clock mechanism, independently of transcription and translation. The peroxiredoxin proteins, noted across eukaryotes and Archaea, produce an autonomous redox rhythm and suggest a mechanism to deal with an increasingly oxidizing environment billions of years ago (Edgar et al., 2012).

NAD+ also serves a critical role in reciprocal regulation of circadian clocks by metabolic status. Sirtuin1 (SIRT1), an NAD+-dependent deacetylase, regulates clock function through deacetylation of PER2 (Asher et al., 2008), while NAD+ levels themselves are intimately tied to mitochondrial function and dictates AMP and ATP concentrations. Steady-state ATP concentrations have been shown to oscillate in multiple tissues (Yamazaki et al., 1994), which regulates AMP-dependent protein kinase (AMPK) activity. AMPK also impacts clock function by phosphorylating CRY (Lamia et al., 2009), and can itself control NAMPT expression (Cantó et al., 2009). Other metabolite-responsive effectors include SIRT6, peroxisome proliferator-activated receptors (PPARs), and glycogen synthase kinase 3 beta (GSK3β), among others, which

collectively comprise a diverse set of sensors to reciprocally impact the core clock (reviewed in detail by Oosterman et al., 2015). It follows then that both diet and temporal feeding patterns may have a profound impact on circadian rhythms. Animals fed a high fat diet (HFD) *ad libitum* display dampened rhythms in core clock and metabolic genes, perturbed temporal feeding rhythms, and lengthened free-running activity periods under constant darkness (Gill et al., 2015, Hatori et al., 2012). Interestingly, limiting the window of HFD feeding, thus enforcing a feeding pattern, restored physiological rhythms and improved metabolic outcomes. This time-restricted feeding (tRF) paradigm has proved salubrious even when switching feeding to the normal resting phase, and is sufficient to restore rhythms of many liver transcripts in *Cry1/2* knockout mice (Vollmers et al. 2009). While the phenotypes are convincing, the mechanism behind tRF is largely unclear. Future studies need to consider why certain times of feeding seem more healthful than others, how nutrients are processed differently vis-à-vis the clock, and whether the metabolic benefits are simply a result of interspersed fasting periods. One emerging consensus from these experiments is that while altered feeding rhythms can reset circadian clocks in the peripheral tissues, particularly metabolically active tissues and nutrient-sensing regions of the brain, the SCN remains largely unaffected by food and ultimately results in organismal desynchrony (Iwanaga et al., 2005). Whether or not chronically desynchronized, but consistent, light and feeding rhythms can still promote health remains to be rigorously demonstrated (Katewa et al., 2016), which is particularly important when considering the proper management of altered sleep schedules, e.g. shift work.

The selective advantage of circadian clocks, particularly for metabolic homeostasis, resides in anticipation (Edery, 2000). To best allocate nutrients effectively, organisms must prepare for anticipated bouts of feeding and fasting during active and rest phases respectively. Reflected in both metabolite and enzyme analyses, glucose synthesis and release from glycogen stores increase at the onset of the active phase, coupled with rhythmic insulin secretion and pancreatic function (Bolli et al., 1984, Kida et al., 1980, Peschke and Peschke, 1998). Digestive processes also increase at the beginning of the active phase, including lipid uptake, enzyme

upregulation, gastrointestinal motility, and gastric blood flow (Pan and Hussain 2007, Kumar et al., 1986). Conversely, organisms prepare for a fasting period by switching from oxidative metabolism to nutrient storage towards the end of the active phase (la Fleur et al., 2001), which may partly explain findings that postprandial glucose levels are higher after dinner than breakfast (Jakubowicz et al., 2013). These sweeping shifts in major metabolic processes have begun to paint a picture of how organisms adapt to their environment and food availability, and may address some uncertainties in the mechanisms of tRF. Amidst the complex interplay of clocks, feeding, and behavior, perhaps the most important role the clock can play in metabolism is to temporally separate anabolic and catabolic processes, hypothesized as a mechanism to maintain metabolic efficiency (Bass, 2012). While logical, this hypothesis needs further experimentation through subcellular metabolite analysis and primary metabolic measurements, including metabolic flux (McGinnis and Young, 2016). Our understanding of circadian metabolism has grown tremendously in recent years, although secondary metabolic measures such as steady state transcript or metabolite levels are inadequate to fully capture these dynamic systems. Additional work is required to truly understand the pathological phenotypes that arise out of perturbed circadian and sleep cycles, and to harness the ubiquitous nature of the clock in a therapeutic manner.

Pathologies, modern lifestyles, and therapeutic opportunities

A temporal distribution in disease symptoms has been observed for hundreds of years, dating back to daily fluctuations of asthmatic symptoms in the medieval times (Lemmer, 2009). More recently, diurnal variation is known to exist for inflammatory response, infection susceptibility, stroke, and myocardial infarctions (Kalsbeek et al., 2012, Reilly et al., 2007). These patterns largely stem from normal clock functioning, considering the rhythms noted in immune cells, circulating cytokines, blood pressure, and heart rate. Conversely, perturbing the clock system has been implicated in numerous pathologies. Clock disruption has been reciprocally implicated as both a driver and a consequence of tumorigenesis (Altman et al., 2015, Masri et al.,

2016). Additionally, the amplitude of circadian cycles dampens with age and is associated with increased activity at night and fragmented sleep in older individuals (Van Someren et al., 1997). These findings have been corroborated by an accelerated aging phenotype in *Bmal* knockout mice, although clock-specific versus clock-independent roles of circadian genes are difficult to ascertain with one animal model (Kondratov et al., 2006). Age-dependent declines in clocks and physiology may be additionally fueled by increased reactive oxygen species (ROS), which build up in greater quantities as sleep becomes increasingly fragmented and redox balance mechanisms decay (Inoué et al., 1995).

Perhaps more alarming and consequential to human health is the ubiquity of circadian and sleep disruption in modern society. Around the beginning of the 19th century, affordable artificial lights ushered in a shift in human lifestyles (Fouquet and Pearson, 2006). Light exposure is now common around the clock, and social interactions can act as a non-photic timing cue at any given moment through social media platforms (Wyse et al., 2014). In addition to increased light-at-night, many human activities today are based indoors. Artificial light does not recapitulate natural sunlight, leaving us with with brighter nights and dimmer days, further dampening circadian clocks. Coupled with longer work hours and/or shift work, humans are largely desynchronized with sunrise and sunset, particularly in urban settings. The recent National Health Interview Survey reported roughly 30% of the American workforce obtain 6 hours of sleep or less per night, with roughly 20% of workers engaging in some form of shift work (Luckhaupt et al., 2010). Shift workers tend exhibit lower sleep quality and quantity (Esquirol et al., 2011, Jay et al., 2006), and are likely to receive less sun exposure than day workers (Puttonen et al., 2010). While causal links are not established, shift workers are also prone to fragmented feeding patterns and less healthy eating, which likely contributes to correlations with metabolic syndrome and obesity beyond already established connections of sleep deprivation and impaired glucose tolerance (Scheer et al., 2009, Esquirol et al., 2011). While constant food availability has generally been viewed as a boon for those with atypical work schedules, the aforementioned tRF studies highlight the need to reconsider eating schedules. Adiposity, bodyweight, glucose, and lipid

metabolism are the most common alterations in a recent meta-analysis of animal studies which model feeding and sleep disruptions (Opperhuizen et al., 2015). Additionally, not all shift workers develop sleep disorders, but such pathologies do exist, such as delayed and advanced sleep phase disorders (Ando et al., 2002, Regestein and Monk, 1995), with relatively little information on metabolic consequences of phase-shifted circadian clocks. More detailed mechanisms of nutrient processing are required in the burgeoning fields of chrononutrition and chronopharmacology to mitigate the metabolic detriments in shift work and habitual sleep disruption.

1.2 *Drosophila melanogaster* as a model in chronometabolic studies

Clock mechanisms in *Drosophila*

The transcription-translation feedback loop in *Drosophila* exhibits similar features to the mammalian clock, and has served as a vital model to delineate molecular mechanisms of clock structures (reviewed in Hardin, 2011, Figure 1.1). The CLOCK/CYCLE (CLK/CYC) complex binds to E-box elements to drive expression of many gene products, including *period* and *timeless*. These protein products (PER and TIM) dimerize and translocate to the nucleus to repress CLK/CYC activity. Phosphorylation of PER/TIM by GSK3B (known as shaggy in *Drosophila*) and CASEIN KINASE 2 (CK2) promotes translocation back to the nucleus, however phosphorylation of PER will by DOUBLETIME (DBT) promotes its binding to the E3 ubiquitin ligase SUPERNUMERARY LIMBS (SLIMB), which leads to ubiquitination and proteolysis by the proteasome, breaking the repressive arm of the clock. Although the molecular mechanisms of the clock are similar, the two-stage mammalian process of SCN entrainment, followed by peripheral clock entrainment, operates differently in flies. Most *Drosophila* cells which contain an oscillator can either directly detect light or receive inputs from photoreceptor cells, demonstrated by light-entrainment of isolated wings and antennae (Plautz et al., 1997). Light rapidly resets the clock in flies through the CRY photoreceptor, resulting in TIM degradation within 30 minutes of exposure (Hunter-Ensor et al., 1996, Myers et al., 1996).

150 cells in the fly brain contain a molecular clock, and some of these are known to control locomotor activity rhythms (Taghert and Shafer, 2006). These cells are clustered into lateral and dorsal neurons (LNs and DNs), of which the small ventral LNs (s-LNv) are necessary for sustained locomotor activity in constant darkness (Helfrich-Förster, 1998). This central oscillator has been shown to modulate the phase of oenocyte clocks, which regulate pheromone production and lipid mobilization, and prothoracic gland clocks, which drive eclosion (Krupp et al., 2008, Myers et al., 2003). However, clocks in the Malpighian tubules, which operates much like the kidneys, and the fat body, exhibiting functions akin to the liver and adipose, can entrain to light and food respectively, independent of the central clock (Giebultowicz 2000, Xu et al., 2011). Thus, while the hierarchy of the clock in flies differs from mammals, circadian clocks ultimately drive rhythms in most major physiological processes, including eclosion, olfaction, courtship, locomotion, and sleep.

Drosophila **in sleep research**

Much like mammalian characteristics of sleep, flies display periods of behavioral quiescence, sleep rebound following deprivation, and increased arousal thresholds (Hendricks et al., 2000). Flies are diurnal, sleeping during darkness, and sleep itself is defined by periods of quiescence lasting at least five minutes, and can enter deeper sleep phases lasting more than 15 minutes with electrophysiological patterns analogous to slow wave sleep in mammals (van Alphen et al., 2013). Given these similarities, and the rapidity and relative accessibility of genetic manipulation, *Drosophila* are a mainstay in sleep genetics research. Sleep/wake cycles form as a complex integration of inputs, including hormones, neuropeptides, neurotransmitters, and physiological perturbations such as feeding or starvation (Keene et al., 2010). Like mammals, sleep homeostasis in flies largely stems from a balance of the monoamines dopamine, serotonin, and octopamine (the invertebrate analogue of norepinephrine) (Livingstone and Tempel, 1983). *Drosophila* genetics has augmented our understanding in the role of these monoamines to regulate sleep and behavior. For instance, impaired dopamine reuptake in the synapse via a

mutated dopamine transporter results in short-sleeping and hyperactive flies (Kume et al., 2005). These *fumin* mutants display similar behavior to increased dopamine signaling via cocaine or methamphetamine, and are countered by inhibitors in dopamine biosynthesis pathways to increase sleep duration (Andretic et al., 2005). Dopaminergic pathways in the fly brain are under control of light and circadian clocks through regulation of inhibitory dopamine receptors in the LNs (Shang et al., 2011). Additional genetic manipulations in monoamine signaling have demonstrated interactions with behavioral patterns to regulate learning, feeding, and metabolic processes, corroborating the utility of *Drosophila* models to understand mammalian sleep networks (reviewed in Nall and Sehgal, 2014). Additionally, *Drosophila* offers some insights into the interaction of sleep and circadian processes with aging, as aged flies demonstrate perturbed dopaminergic signaling and increasingly fragmented sleep with age (Kayser et al., 2014). Circadian and aging physiology may also couple with metabolism, as clock mutants lose cycling in ROS and antioxidant-defense enzymes (Krishnan et al., 2008). *Drosophila* may yet provide much insight into the circadian-sleep-metabolic connection as our elucidation of conserved metabolic mechanisms in flies grows.

Metabolism and nutrient response in *Drosophila*

Drosophila were among the earliest model organisms to demonstrate links between metabolism and lifespan (Loeb and Northrop, 1917), and have recently regained interest in dissecting mechanisms of metabolic homeostasis vis-à-vis diabetes and obesity (Leopold and Perrimon, 2007). In addition to the major metabolic organs such as fat bodies, Malpighian tubules, and oenocytes, flies also harbor the pars intercerebralis-corpora cardiac system, which exhibits similar functions to the hypothalamus-pituitary system and pancreas (Bharucha, 2009). Flies digest food through the crop and midgut, analogous to the stomach and intestine (Pitsouli and Perrimon, 2008), although nutrients are then released into the hemolymph as part of an open circulatory system, thus demanding unique mechanisms of nutrient distribution. Flies maintain efficient nutrient transport by harboring a tubular heart structure and circulating trehalose, a

disaccharide of glucose, to effectively increase sugar concentrations while maintaining osmolality (Klowden, 2007). Flies respond to the presence of nutrients through secretion of insulin (commonly referred to as *Drosophila* insulin-like-peptides, or dILPs) for glycogen and triglyceride storage (Barbieri et al., 2003). Unlike mammals, flies express 8 dILPs, which vary in tissue distribution but bind to one ubiquitous insulin-like receptor (Fernandez et al., 1995). dILP6 is most notable, activating insulin-signaling pathways in the fat body to promote fat storage, similar to mammalian systems (Saltiel and Kahn, 2001). Additional controls on dILP release include TOR-dependent sensing of amino acids and trehalose, and Mio/Mondo-dependent sensing of glucose, a homolog to ChREBP (Geminard et al., 2009, Postic et al., 2007). The fat body also contains insulin-responsive AKT pathways (Garofalo 2002). Conversely, the neurosecretory cells of the corpora cardiaca secrete adipokinetic hormone (AKH), which acts much like glucagon to activate fat body cells and promote glycogenolysis, trehalose production, and AMPK-dependent lipolysis (Kim and Rulifson, 2004, Rayne and O'Shea, 1994, Staubli et al., 2002).

Conserved nutrient response mechanisms in flies have recently prompted efforts to model diet-induced pathologies. Both genetic manipulation of dILPs and a high-fat diet can disrupt glucose homeostasis and produce similar phenotypes to metabolic syndrome and diabetes (Zhang et al., 2009, Birse et al., 2010, Musselman et al., 2011). These genetic and dietary manipulations have also yielded cardiomyopathies, which was recently expounded to demonstrate the benefits of tRF on diet-induced cardiac decline (Gill et al., 2015), implying additional synergy of metabolic and circadian processes which can be modeled in *Drosophila*.

Clock and metabolic interactions in *Drosophila*

Circadian research greatly benefited from the advent of accessible and affordable oligonucleotide arrays in the early 2000s. Multiple microarray analyses were performed in both heads and bodies of flies, and later expanded into tissue-specific circadian analyses (Claridge-Chang et al., 2001, McDonald and Rosbash, 2001, Ceriani et al., 2002, Lin et al., 2002, Xu et al., 2011). The complexity of these datasets prompted a meta-analysis to glean overlapping transcript

oscillations, although at the time most of these hits were either part of the core clock or uncharacterized (Keegen et al., 2007). Some annotated metabolic genes emerged as overlaps in at least a subset of these studies, including histidine decarboxylase, aldolase, and UDP-glycoslytransferase. Transcripts implicated in glucose metabolism and transport, as well as neurotransmitter biosynthesis, also emerge across multiple studies, however the time of peak expression is inconsistent. Additionally, fat body transcriptomics depicts an even more complicated picture, where many genes peaked at different times from whole-body analyses (unpublished). Unsurprisingly, the fat body responded to an enforced feeding pattern in the background of a *clk* mutant strain much like liver transcripts in *Cry1/2* knockout mice (Vollmers et al., 2009), highlighting the possibility of conserved mechanisms of feeding entrainment. Although phases differed across studies, multiple enzymes involved in reduced glutathione (GSH) production oscillate, which serve to maintain levels of intracellular antioxidants and counter ROS oscillations (Krishan et al., 2008, Beaver et al., 2012). GSH itself displays a temporal pattern, though not as strongly as the biosynthetic enzymes, which may reflect an intracellular objective to maintain steady-state concentrations. Clock mutant strains display impaired transcript oscillations in GSH enzymes, and may explain increased ROS, peroxidated lipids, and accelerated aging and neurodegeneration in these flies (Krishnan et al., 2009, Krishnan et al., 2012). While focused on aging, a recent review highlights the role of peroxiredoxins in *Drosophila* to combat oxidative stress and maintain ATP levels, creating a vital link to redox cycles which requires further investigation (Orr et al., 2013). With a resurgence of metabolic studies in *Drosophila*, many postulations can be made in the chronometabolic connections, however analyses commensurate with these complex and dynamic systems must be employed to truly understand these phenotypes.

Current methodology in *Drosophila* metabolism studies

Metabolic homeostasis is a complex process in any living system, which is difficult to experimentally ascertain with experimental techniques in genetics, transcriptomics, and hormone

assays. Assessing bulk changes in lipids and sugars can aid interpretation of metabolic response to nutrient intake, which have been employed in *Drosophila* using colorimetric glycogen and triglyceride (TAG) assays (Tennessen et al., 2012). Additional separations of lipid groups, for example fatty acids and diacylglycerols (DAG) can be further measured using thin-layer chromatography (TLC), which is particularly relevant for flies as DAGs comprise the major circulating form of lipids in hemolymph (Hildebrandt et al., 2011). These assays, if done meticulously, can be performed in dissected flies to discern tissue-specific lipid contents. Analysis of sugars is typically confined to glucose and trehalose assays, which require careful enzyme-based assays to separate glucose before and after trehalose digestion (Teleman et al., 2005). Analysis of circulating sugars can be performed through isolation of the hemolymph, although tissue-specific analyses likely fall out of accurate detection (Lee and Park, 2004). While still considered a secondary metabolic analysis, measurements of energetics through steady-state NAD+ or ATP concentrations can provide valuable information to infer the dynamics of fuel oxidation and nutrient-switching. ATP assays do exist, but given chemical instability, has proven difficult to measure accurately (Park et al., 2006).

More sophisticated technologies are required to detect specific sugars, lipids, and other metabolites such as amino acids. Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are common analytical strategies in the burgeoning field of metabolomics. Coupling MS techniques to chromatographic separation, including gas (GC-MS) or liquid (LC-MS), can additionally broaden the scope of metabolite analysis. These approaches have largely remained in the dark in *Drosophila* chronometabolic studies, however metabolomics has been employed in aging and developmental biology (Thuy An et al., 2014, Hoffman et al., 2014, Laye et al., 2015, Tennessen et al., 2014), and tissue-specific analyses has been performed to create spatial maps of the fly metabolome (Chintapalli et al., 2013). Encouragingly, one study utilized NMR in whole flies across circadian time to discern cycles in amino acids and energetic metabolites, offering hope in a new analytical frontier to unveil the reciprocal regulation of clocks and metabolism (Gogna et al., 2015). As these platforms expand in scope, the overwhelming complexity of

rhythms in transcripts, proteins, PTMs, and epigenetics can be uprooted and replanted with a deeper and more accurate phenotype of metabolite dynamics.

1.3 LC-MS metabolomics and multivariate statistical analysis

LC-MS metabolomics methodology

While metabolomics is considered a relatively new member of the high-throughout 'omic platforms, biofluid analysis has been performed since ancient China and the Middle ages to detect high glucose levels or otherwise unusual smells and colors in urine (van der Greef and Smilde, 2005, Nicholson and Lindon, 2008). More recently, the idea of metabolic 'fingerprinting' had been proposed by Roger Williams in the 1940s as a method to identity unique signatures across healthy and diseased individuals using TLC (Williams, 1951). In time, advances in GC and LC ushered in a new era of quantitative metabolite profiling (Horning and Horning, 1971, Pauling et al., 1971). This new wealth of knowledge created new perspectives on metabolism, leading to the first compilations of metabolic networks by Donald Nicholson (Nicholson, 1970), lauded by Nobel laurate Ernst Boris Chain as one of the great achievements in biochemical research (Chain, 1965). Continued developments in analytical chemistry have since enabled detection of thousands more metabolites than thought to exist at the dawn of functional biochemistry, revealing a yet incomplete picture of metabolic networks. Improved metabolomic analyses hold great promise for the fields of food science, drug development, toxicology, and translational medicine, as metabolites are thought to serve as a readout of both the actions of intrinsic enzyme activities and exposures to extrinsic environmental factors.

The detection methods for current metabolomics' technologies include GC-MS, LC-MS, capillary electrophoresis (CE-MS), and NMR. While GC-MS analyses are still employed for analyzing specific subsets of the metabolome, and CE-MS improvements are rapidly expanding the breadth of metabolite analysis (Ramautar et al., 2009), LC-MS currently has the most widespread use and is typically considered to provide the deepest metabolome coverage (Theodoridis et al., 2012). The LC-MS metabolomics workflow considers all aspects of metabolite

detection, from sample preparation to chromatographic separation and detection of metabolites by the mass analyzer.

The first major step, sample extraction, requires careful consideration of the experimental objectives (reviewed in Cajka and Fiehn, 2016). The mass or volume of the biological sample to be extracted must be rationalized given mass spectrometers cannot accurately quantify metabolites across the physiological range of concentrations, which may vary up to eight orders of magnitude. Extraction methods may be tailored to specific classes of metabolites in accordance with the predicted octanol/water partition coefficients, however this approach will likely preclude holistic metabolite profiling. One common tradeoff is to perform a single extraction to recover more metabolite classes simultaneously (e.g. lipophilic and polar metabolites), with the added benefit of speed and simplicity. Single-extraction approaches can include organic solvent-protein precipitation or biphasic liquid-liquid extractions such as the Bligh-Dyer approach, which yields aqueous and organic layers amenable to separate LC-MS methods for small-molecules and lipids (Bligh and Dyer, 1959). Unfortunately, no single method can recover all metabolites effectively, however many options exist to suit the needs of the researcher.

Likewise, chromatographic separations can be tailored to enhance detection of specific metabolites of interest. Direct-infusion MS is common, however chromatography enables the separation of isobaric species (compounds of identical mass) and mitigates ion-suppression, which refers to the competition metabolites face for ionization when simultaneously present in the MS source (Blanksby and Mitchell, 2010). The most common bifurcation of chromatographic methods in metabolomics is reverse-phase (RPLC) and hydrophilic interaction liquid chromatography (HILIC), for analysis of nonpolar and polar metabolites respectively. RPLC is typically achieved through C18 column chemistry, which binds strongly to hydrophobic compounds, and a gradient of polar to nonpolar mobile phases, generating a temporal elution of compounds off the solid phase with increasing hydrophobicity. RPLC traditionally has more widespread use in analytical chemistry than HILIC, and is suitable for lipid profiling (lipidomics). HILIC compound retention mechanisms operate slightly differently in principle, however opposing

conditions generally hold true: solid phases consisting of silica or derivatized silica to retain polar compounds, and mobile phase gradients from organic to aqueous solvents for compound elution off the solid phase. Many variations of these approaches exist, reviewed in Cajka and Fiehn, 2016, including an array of mobile phase additives to enhance metabolite ionization, and stationary phases of smaller particle sizes to improve chromatographic separation of compounds with similar physiochemical properties. Workflows which utilize both RPLC and HILIC greatly expand metabolome coverage, albeit at additional cost and time to perform multiple analyses.

MS detection adds another dimension of complexity to LC-MS workflows. For detection on a mass spectrometer, metabolites must exist as a charged species through ionization in the MS source. Multiple options exist for ionization, most classically through election impact (EI), however for smaller and labile molecules, softer ionization techniques such as electrospray ionization (ESI) are now common, which may operate to positively or negatively ionize molecules (ESI+ or ESI- modes respectively). MS metabolomic analyses are conceptually divided into either targeted or untargeted approaches, where targeted approaches aim to accurately detect specific metabolites with greater specificity and sensitivity, and untargeted approaches aim to detect as many metabolites as possible, unbiasedly and without quantification. Targeted approaches typically employ mass spectrometers capable of fragmentation (MS/MS), such as triple quadrupole mass spectrometers (QqQ) or quadrupole/linear ion trap instruments. While mass resolution is low on these instruments, specificity can be achieved through MS/MS by analyzing the parent mass-to-charge ratio (m/z), followed by the daughter m/z(s) after imparting energy on the molecule in the collision cell. This parent-to-daughter mass transition is commonly called multiple reaction monitoring (MRM), and is the primary means to detect and quantify metabolites in these instruments (Figure 1.2). With faster scan speeds and ion-polarity switching, newer instruments can scan for hundreds of MRMs, increasing the scope of targeted metabolomic analyses (Rhoades and Weljie, 2016, Yuan et al., 2012). Time-of-flights (TOFs), and other high-resolution instruments such as Orbitraps, are most commonly used for untargeted metabolomics to detect thousands of metabolic features. While powerful, these instruments on

their own are typically reserved for exploratory analyses and yield many m/z values without metabolite identification. Extensive follow-up work, through either MS/MS or bioinformatics, is required to identify m/z values as known metabolites. Some instruments combine high-resolution and fragmentation analysis to collectively achieve selectivity, sensitivity, and mass accuracy. As databases continue to improve metabolite identification from high-resolution MS metabolomics datasets, the utility of untargeted metabolomics to probe biology is expected to grow.

Despite recent advancements, metabolomics invariably encounters a high diversity of physiochemical properties across metabolite classes. Unlike genomics, where all genes can be measured, metabolomics has yet to delineate the entire metabolome, which may contain tens of thousands of metabolites (Wishart et al., 2013). Paired with the wide range of metabolite concentrations found in biological samples, much work remains to develop equivalently diverse and accurate analytical tools, in addition to building new computational workflows for high-throughput metabolite identification. Method development appropriated for multivariate detection, coupled with multifactorial design principles (e.g. choice of mobile phase, LC gradient, MS ionization parameters, etc.), generate a combinatorial complexity in optimization objectives. Commensurate multivariate workflows, such as design of experiments (DoE), can guide the analytical chemist to efficiently design complementary metabolomics methods with high metabolome coverage, sensitivity, and accuracy (Gika et al., 2012).

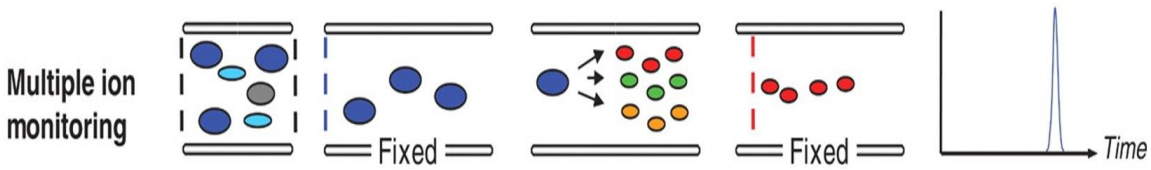


Figure 1.2 Mechanism of multiple reaction monitoring (MRM) and representation of an LC-MS/MS chromatogram. Reprinted with permission from Domon and Aebersold, 2006.

Design principles and statistical approaches to process and analyze metabolomic data

Large-scale metabolomic analyses must consider data acquisition, processing, and analysis to obtain meaningful biological inference. While standardized workflows in LC-MS metabolomics are lacking, representative procedures reviewed by Dunn and colleagues are most common (Dunn et al., 2011). After sample extraction, samples are analyzed via LC-MS in a randomized manner to avoid biases due to analytical drift, referring to the variance exhibited by LC-MS methods across days or weeks of use. Quality control samples (QCs) are interspersed in the run order, which can contain either a mix of metabolite standards, or consist of a pooled sample of all biological samples under study. QC injections can then be used to inspect both analytical drift and response variance. Drift can be corrected by regression fitting to the QC data points, metabolite by metabolite. Additionally, any metabolites which display high variance, which may arise if the signal-to-noise (S/N) is low, can be dropped from the dataset. Stringency in data correction and filtering may depend on the objectives of the experiment, for example exploratory preclinical analyses versus targeted and quantitative clinical analyses.

After data processing, multivariate or univariate methods may be used to discover differences across experimental groups (Kotlowska, 2014). Univariate methods, such as t-tests, may have utility in biomarker research, however given the dimensionality of metabolomics data, must consider multiple testing correction, such as Bonferroni or false discovery rate adjustments (Dunn, 1961, Benjamini and Hochberg, 1995). In reality, phenotypes are driven by relationships of multiple variables. Two of the most common multivariate approaches in metabolomics are principal components analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) (Trygg et al., 2007). PCA is an unsupervised approach which seeks the maximum variation among the metabolites (X variables) to derive new latent variables, or principal components. PCA yields global structures in metabolomic data irrespective of biological groups, and is often used as a preliminary analysis. OPLS-DA is a supervised approach, which aims to derive latent variables from maximal covariance of the metabolites with biological groups (Y variables). Metabolites which contribute to the largest group separation can then be used in a

predictive manner for diagnostic or biomarker purposes (Mapstone et al., 2014), or overlaid onto pathways or networks to understand metabolic mechanisms. Other approaches may be used, including clustering analysis or Bayesian-based methods, however the ultimate objectives in analyzing metabolomic datasets remain similar.

LC-MS metabolomics in chronobiology studies

Currently, there is a dearth of metabolomic data for sleep and circadian analyses in *Drosophila* (Gogna et al., 2015), however metabolomics is quickly being adopted in other model organisms and humans, recently reviewed in Brown, 2016, and Rhoades et al., 2017 (Figure 1.3). Rhoades et al. compiled a meta-analysis of recent human studies, revealing overlapping metabolite hits across multiple chromatographic and MS detection methods. These hits include phosphatidylcholines, medium and long-chain acylcarnitines, and amino acids. Members of these compound classes have been shown to oscillate in murine liver and adipose, with dependency on a fully functional genetic clock (Eckel-Mahan et al., 2012, Castro et al., 2015). More causal links of the clock to targeted metabolic outputs in mice have also been discovered, including unique signaling roles of a specific lipid (PC 18:0/18:1) to regulate metabolic oscillations across tissues (Liu et al., 2013), and functional roles for polyamines to maintain robust clock function during aging (Zwighaft et al., 2015). The adipose clock has been shown to impact feeding behavior through rhythmic release of lipids in a feedback mechanism to the hypothalamus (Paschos et al., 2012). Tissue-specific clock manipulations in genetic mouse models have simultaneously uncovered new connections between clocks and metabolism across tissues, while opening a Pandora's box of complicated mechanisms to maintain systemic metabolic homeostasis across circadian time. Additional experimentation will further elucidate our understanding of this complexity, however new approaches are required to derive interpretable phenotypes at an organismal scale. Appropriate animal models are also necessary to advance this field, as *in vitro* models of circadian rhythms do not faithfully recapitulate the complex interactions of environmental exposures, feeding, behavior, and tissue-specific metabolism. Metabolomics holds

promise to uncover the true depth of rhythmic metabolic outputs, as the enrichment of rhythms in transcripts and their median fold-change in murine liver is 1.51 and 18.3%, respectively (Hughes et al., 2009), compared to 1.98 and 50% in detectable metabolites under the same experimental protocol (Krishnaiah et al., 2017). Overlapping metabolite rhythms to mice and humans have yet to be demonstrated in *Drosophila*, but given conserved mechanisms of sleep, circadian rhythms, and nutrient-sensing pathways, many opportunities exist to further explore these chronometabolic connections.

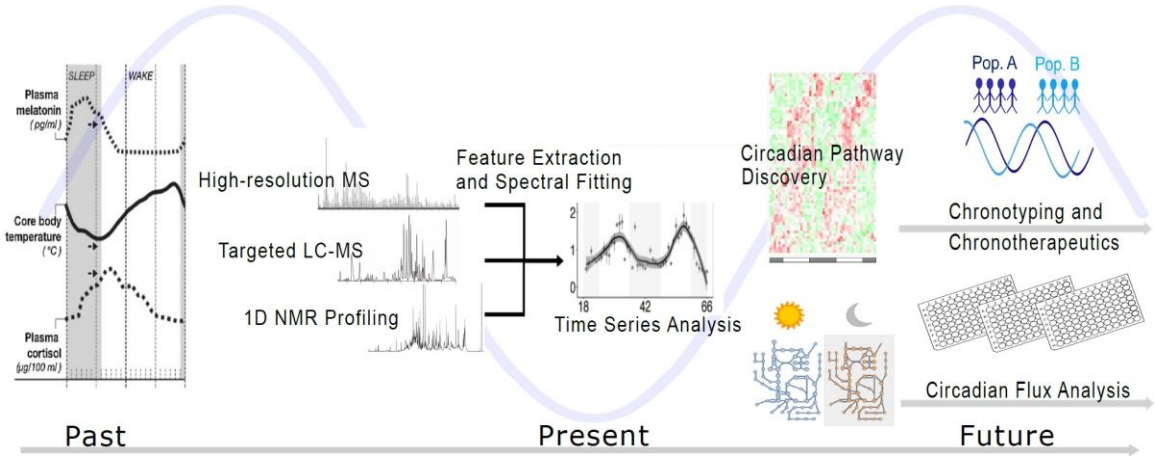


Figure 1.3 Timeline of metabolomics in chronometabolic studies, both current approaches and future directions. From Rhoades et al., 2017.

Statistical tests to discern periodicity in data

Assessing rhythmic patterns in time-series data requires appropriate statistical tests. True oscillations, resembling a sine wave, can be tested with rank-based correlations of monotonic sequences, most frequently performed in the circadian field using JTK_CYCLE (Hughes et al., 2010). Wavelet tests provide additional information relevant for circadian rhythms, including the peak of oscillation (phase) and height (amplitude), which can directly answer both when a metabolic process is most active and the relative strength of the rhythm. Experimental design plays a critical role in deciphering periodicity with statistical significance. Ideally, sampling frequency would be high such that a wave can be clearly formed across time, however is rare in practice due to cost and resource constraints. Reasonable compromises of time resolution and resource allocation include sampling every two or four hours, which is sufficient to test across 24-hour cycles. However, shorter sampling times would be required for ultradian rhythm analysis, as 4-hour sampling resolution would be insufficient to find 8-hour cycles. Recent work has shown increased resolution also provides increased statistical power, as true positive rates were dramatically higher in one-hour sampling compared to two or four hours (Krishnaiah et al., 2017). JTK_CYCLE is commonly executed such that waves are either fit to replicates within one day of sampling, or replicates spread across multiple days. Care must be taken in testing rhythmicity to mitigate day-to-day biases, which may be mitigated with data permutations. Additionally, monotonic changes across time, for example metabolic shifts in concert with changes in cell culture media, must be separated from circadian-dependent temporal patterns (Sengupta et al., 2016). Alternative approaches such as time-dependent analysis of variance (ANOVA) or pairwise univariate testing may also be necessary as diurnal metabolic outputs may not exhibit classic waveform patterns.

1.4 Application of stable isotope tracers in metabolism research

Definitions and usages of stable isotopes

Isotopes, most simply, refer to forms of the same element with different numbers of neutrons. The most commonly used element in isotope-based metabolic studies is carbon, with ^{12}C comprising 98.9% of all carbon on Earth, and the heavier ^{13}C carbon comprising the roughly remaining 1.1%. ^{14}C carbon also exists, and while it is unstable and radioactive, serves important purposes in metabolic studies. In the middle of the 20th century, radioisotopes, including ^{14}C, were frequently utilized to trace metabolic networks, which Nobel laurate Ernst Boris Chain describes as resulting in 'an almost explosive expansion of the field of the study of metabolic pathways' (Chain, 1965). At the time, radioisotopes were effective given their ease of detection, but were largely nonspecific for probing larger swaths of metabolism. Today, heavy isotopes still pervade metabolic research, with stable ^{13}C tracers recently entering metabolomics for pathway discovery on the heels of improved MS and NMR instrumentation (Giavalisco et al., 2008). Stable isotopes are also used for quantification of endogenous metabolites on GC and LC-MS platforms, as they are easily identifiable by an atomic mass unit (amu) shift of 1.00335 for each ^{13}C atom. Metabolites may vary many orders of magnitude and suffer from ion suppression by coeluting species off the chromatographic column (Shi, 2003), therefore to obtain accurate quantitation, metabolites may be analyzed with a known concentration of an internal standard (most commonly a stable isotope form of the compound of interest). In theory, any ion suppression effects on the endogenous metabolite will also apply to the internal standard, thereby yielding a reference compound for accurate quantification.

A renaissance in metabolic tracers has spawned a new interest in metabolic flux analysis. While metabolomics itself adds a layer of understanding to biological systems, static profiling of metabolite levels does not provide dynamics of metabolite flow. For instance, if a metabolite concentration is high, one cannot be certain of either an increased flux from a producing enzyme, or decreased flux through a consuming reaction (Creek et al., 2012). Additionally, if one were to think of a metabolite as a node in a complex graph network of metabolism, the concentration alone would be an obfuscation of numerous metabolic processes. Fluxes are thought to be the

most fundamental assessment of metabolic phenotypes, and is the focus of intense research efforts for both basic biology and therapeutic discovery applications.

Applications of stable isotopes in intracellular flux analysis

Isotope-tracer studies are most commonly performed *in vitro*, given uptake and effluxes of labeled metabolites are more accessible than *in vivo* (Figure 1.4). These experiments are suitable for comparative analysis of metabolic activity, for example healthy cell lines to genetic knockdowns or cancer (Fan et al., 2013). Although the flux of atoms through enzymes cannot be directly measured per se, the detection of isotope incorporation into metabolites downstream from a known input tracer is a viable alternative. Through repeated sampling of the cell culture media, the rates of change in isotopically labeled metabolites can be calculated and comprises the most basic form of metabolic flux modeling. The addition of intracellular isotope analysis adds significant biological insight, albeit with increased complexity (Sauer, 2006). $^{13}C_6$ glucose, referring to all six carbons of glucose consisting of ^{13}C (also called uniformly labeled glucose), is one common tracer for flux analysis through central energy metabolism, although other tracers may be used to probe secondary metabolism (Winder et al., 2011). Labeled metabolites, called isotopologues, are typically measured on GC or LC-MS to generate the mass distribution vector (MDV). The MDV describes the relative abundance of a given isotopologue, from zero to n labeled carbons. For instance, a 0.10 enrichment of an M+1 isotopologue implies ten percent of the total metabolite signal consists of that metabolite with one labeled carbon. This approach considers parent m/z values without positional information of the labeled carbon, which would require MS/MS or NMR. Isotopologues which differ in the position of the ^{13}C atom(s) are called isotopomers, and can provide considerably greater reaction specificity over isotopologues alone. Inevitably, analytical detection of full isotopomer distributions with larger molecules becomes unwieldy as each parent isotopologue contains $\binom{n}{k}$ isotopomers (Buescher et al., 2015), which is nontrivial in MS/MS methods but has been performed with aspartate (Choi et al., 2012). Before drawing biological inference from MDVs, data processing measures must be taken to correct for

the abundance of natural isotopes, including the 1.1% natural abundance of ^{13}C. For more than one labeled carbon, these calculations become increasingly complex. Fortunately, isotope-correction procedures have been made available and are part of routine metabolic flux workflows (Moseley, 2010).

Stationary flux analysis underlies most published flux analyses, and refers to the assumption that upon measurement, the isotope enrichment for a given metabolite is in equilibrium. This assumption holds reasonably well for pathways with high flux and turnover, notably glycolysis, which can reach steady-state labeling within minutes (Buescher et al., 2015). Isotopic steady-state is independent of the metabolite levels themselves, meaning MDV measurements are adequate to gauge pathway activities in a relative manner (Zamboni et al., 2009). If the pathways contributing to an MDV are known, relative comparisons may be made even in situations of non-steady state, for example in amino acid metabolism, which may take days to reach isotopic equilibrium (Buescher et al., 2015). However, if pool sizes of a metabolite are dramatically different across biological conditions, the MDVs alone may yield misleading information. For instance, if a given isotopologue enrichment is one percent in condition A and ten percent in condition B, but the pool size of that metabolite is ten-fold in condition A, then the flux of atoms into that metabolite pool is roughly equivalent, even if the MDV differs. Quantifying the metabolite concentration and MDV at high temporal resolution before and after reaching isotopic equilibrium provides the most accurate estimation of flux, however is difficult experimentally, analytically, and computationally.

Even if isotopic equilibrium is not attained, the metabolite pool sizes themselves must remain in steady-state to estimate metabolic flux (Zamboni et al., 2005, Weitzel et al., 2013). If this assumption does not hold, nonstationary flux analysis must be considered, reviewed in detail by Wiechert and Nöh, 2013. These analyses are additionally complex and currently out of reach for mammalian systems. However, the rigor required for tracer studies can be largely dictated by the biological question, as relative pathway activities may be adequate to discern metabolic differences across experimental groups. Of note, any given isotopologue likely derives from

multiple sources, especially in secondary metabolism, and may require positional carbon information to attribute labeling patterns to specific reaction activities. Some of these concerns of reaction specificity may be mitigated by utilizing multiple tracers, for example combinations of glucose, palmitate, and amino acids. Recent advances in metabolomics data processing algorithms can abstract isotope enrichments in high-resolution untargeted LC-MS datasets, opening new possibilities for discovery of active metabolic networks (Huang et al., 2014, Capellades et al., 2016). While great advancements have been made in the metabolic flux field, much work remains to rigorously derive quantitative flux measurements and understand complex labeling patterns. Vis-à-vis uncertainties in the interpretation of isotopologue sources, future work will need to focus on building reaction networks for isotope labeling patterns with atom-level resolution. While atom-mapping reactions exist, such as in the MetaCyc database (Caspi et al., 2016), there are currently no methods to map isotopologues onto models of metabolism at a genome scale (Chokkathukalam et al., 2014).

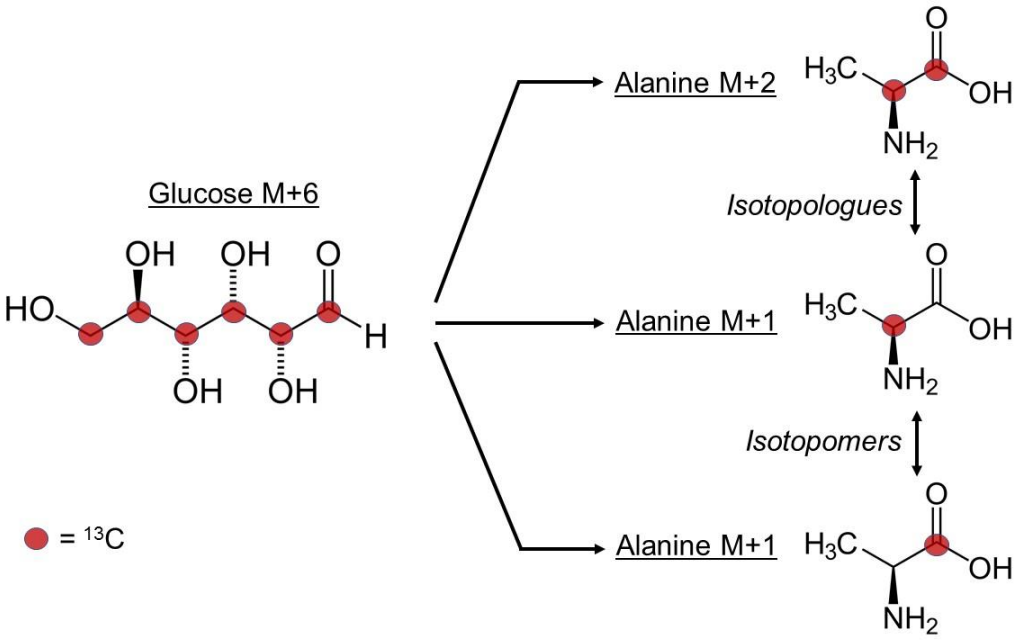


Figure 1.4 Depiction of isotopologues and isotopomers, exemplified by $^{13}C_6$ glucose to isotope enrichments in alanine.

Stable isotopes *in vivo*

Stable isotopes serve important roles in animal and human research, with the latter explained in detail by Wolfe and Chinkes, 2005. Fan et al. utilized a bolus injection of $^{13}C_6$ glucose in a mouse model of lung cancer as a proof-of-principle to track relative changes in isotopologue patterns across multiple tissues (Fan et al., 2011). However, repeated sampling from mice is difficult (Castro-Perez et al., 2011), and otherwise would require many mice to generate a sufficient time-series of isotopologue patterns to calculate reaction rates. Justification of isotope tracing in animals is questionable for exploratory analyses given the limited interpretability of these datasets. Conversely, for human studies, considerable effort has been made to correctly model reaction kinetics for well-defined pathways. Infusions are the preferred mechanism of isotope delivery, as obtaining isotopic equilibrium is easier than through injection boluses. (Suh et al., 2003). In situations where physiological steady-state is perturbed, such as feeding or exercise paradigms, isotopic steady-state will no longer hold, and additional calculations of isotopologue rates of appearance and disappearance must be considered (Steele, 1959). These calculations generally consider the tracer to exist in a single compartment, such as blood, with instantaneous and homogenous mixing. Single-compartment models are useful, however may falter as tissues metabolize and resynthesize nutrients to be released back into the bloodstream. Models which consider multiple compartments do exist, but require repeated biopsies from each considered tissue with sufficient temporal resolution.

Despite tremendous advancements in modern MS and NMR metabolomics instrumentation, the detection of stable isotopes *in vivo* requires significantly higher concentrations than radioactive isotopes. Large doses of stable isotopes may perturb physiological steady-state, yielding a state of isotopic disequilibrium. Additionally, compartmental models may yield spurious information if physiological perturbations alter the tracer's volume of distribution. Regardless, advancements in isotope detection and reaction modeling have provided valuable information for carbohydrate, lipid, and amino acid oxidation. Beyond glucose, palmitate and glycerol isotopes have uncovered lipid oxidation rates (Greenough et al., 1969), and

phenylalanine, tyrosine, and alanine have analyzed both amino acid catabolism and protein anabolism rates (Engelen et al., 2012, Hartl et al., 1990). While some assumptions in *in vivo* tracer kinetic modeling may not hold, they serve reasonably well to discern real metabolic phenotypes.

Applications of isotope tracers in chronobiology to date

Isotope tracer studies have widespread usage *in vitro* for cancer metabolism research, and in humans to gauge metabolic responses to exercise and feeding, however few studies exist which leverage stable isotopes in chronobiology. Metabolism of isotopically labeled glutamate has been shown to depend on the relative levels of gamma-aminobutyric acid (GABA) and glutamine in astrocytes and neurons, suggesting control of sleep/wake cycles through a metabolic mechanism (Qu et al., 2001, McKenna and Sonnewald, 2005). Pharmacological inhibition of GABA transaminase, which alters GABA and glutamate levels, results in a perturbed metabolic homeostasis of the TCA cycle and energetic metabolites (Maguire et al., 2015). This effect derives in part through increased transfer of glutamate carbons to lactate, demonstrated via *in vitro* labeling analyses of cultured *Drosophila* neurons fed $^{13}C_5$ glutamate, and likely arises from compensation for impaired GABA metabolism.

One recent study reported impaired lipid synthesis in *timeless* mutants, measured through ^{14}C incorporation into bulk lipids from ^{14}C-glucose added to fly media (Katewa et al., 2016). This effect is only visible under a tRF paradigm, bringing into question the clock-dependent versus independent roles of *timeless*, as this effect was not tested in other clock mutants. Isotope-feeding studies have also been performed in flies on high-sugar diets to understand dietary influences on fat body metabolism (Musselman et al., 2013). Ancillary to the study objectives, isotope enrichments in fatty acids, triglycerides, and trehalose reached an isotopic equilibrium roughly 24 hours after transfer off the food containing $^{13}C_6$ glucose. This observation is important in designing future labeling studies. Given sugars and lipids are known to oscillate with 24 hour periods (Scheer et al., 2009, la Fleur et al., 1999, Dallmann et al., 2012),

flux modeling on shorter timer scales would require nonstationary flux analysis, which is currently infeasible in whole organisms (Weichert and Nöh, 2013). Qualitative comparisons of isotope enrichments may be adequate to discern cycles in metabolic flux, however acute labeling experiments would be required to attain appreciate sampling resolution for rhythmicity tests. Forcing flies to acutely consume a large meal of a labeled nutrient would likely require starvation protocols, which may confound results considering perturbing the activator and repressive arms of the clock can have opposing effects on the starvation response (Xu et al., 2008), however acute feeding through e.g. an injection bolus may circumvent this concern.

Diurnal nutrient adsorption of glucose and small peptides have been observed at light:dark transitions in wild-type and *clock* mutant mice through an *in situ* loop mechanism in the jejunum (Pan and Hussain, 2009). Decay from the ^{14}C tracers were detected in the portal vein one hour after infusion to discover increased adsorption of both sugars and peptides in wild-type mice at the onset of the active phase. Temporal patterns in adsorption were lost in *clock* mutants, which also demonstrated an overall increased and decreased adsorption of sugars and peptides respectively. This study marks an important step forward to demonstrate true nutrient processing changes across time under control of the circadian clock, beyond secondary predictions made from expression or enzyme abundance. Future studies should continue to expand the scope of isotope tracing in defining the dynamics of circadian metabolism. As previously mentioned, many oscillating metabolites from previous metabolomics analyses derive from lipid and amino acid metabolism. Tracing experiments should be appropriated to corroborate these hits by probing relative activities of relevant pathways, even if non-steady state metabolic conditions prevent quantitative metabolic flux *in vivo*. Additionally, meaningful inferences must be made from labeling patterns in these secondary metabolic processes, which will require expanded metabolic models and computational approaches which operate at a genome scale.

New era of genome-scale networks and *in silico* metabolic flux

With the advent of whole-genome sequencing, comprehensive metabolic networks can be constructed to consider all metabolites and reactions in an organism (Feist et al., 2009). This compendium of reactions can be converted to a genome-scale stoichiometric matrix, mathematically considering all reactants and products. These network reconstruction efforts underlie the field of constraint-based modeling (CBM) (reviewed in Bordbar et al., 2014), which differs from inference-based approaches to build networks *de novo* from patterns in experimental data (reviewed in Bonneau, 2008). CBM is a computational approach to understand dynamics of metabolism, signaling, and transcriptional regulation by imposing constraints on the metabolic network and predicting a set of possible flux values for each reaction. Constraints may include thermodynamics, diffusion limits, and 'omic datasets to improve flux predictions (Weisz, 1973, Shlomi et al., 2008). Thus, a whole genome-scale metabolic model for any given organism can serve as the scaffold upon which context-specific models can be generated, such as tissue-specific networks.

CBM is currently on the cusp of wider usage in biomedical research, and has been used for drug target discovery and the exploration of new reaction mechanisms (Frezza et al., 2011, Fischer and Sauer, 2003). One appealing development in genome-scale flux predictions is the objective to restore the flux vector in aberrant metabolism to a healthy state through knockout simulations (Yizhak et al., 2013). Additionally, individualized human kinetic models combine enzyme kinetics and personalized 'omic datasets, which hold promise for tailoring personalized medicines to control metabolism (Bordbar et al., 2015). Genome-scale models continue to improve in accuracy and annotation, and our current understanding of canonical biochemistry pathways from mid-20th century research is already in flux as metabolic simulations continue to expand further from the hub of central carbon metabolism (Sauer, 2006). With the rapid expansion of computing power in the 21st century, probing metabolic networks through complex experimental designs, such as isotope tracer approaches, will become increasingly feasible and interpretable. One current design feature missing from CBM simulations is tracking individual carbon fates in metabolic networks, however metabolic simulations at atom-resolution are within

reach, and are expected to bring science closer to unearthing the complete cellular metabolome once thought to be solved in Nicholson's early metabolic maps (Nicholson, 1970).

1.5. Overview

The associations of sleep and circadian disruption with poor metabolic outcomes warrant further investigations of causality. The exploration of translatable biological mechanisms hinges on the appropriate applications of cutting-edge technologies across model organisms. Powerful genetic tools and systems-level discovery platforms have provided tremendous insight into the basic mechanisms of clock and sleep processes in *Drosophila*, however chronometabolic connections remain largely unexplored. *Drosophila* provide a tractable model to separate environmental and endogenous clock-induced metabolomic changes, which cannot be adequately modeled *in vitro*. An underlying extension of this metabolomics approach, irrespective of model organism, is the hypothesis that steady-state cycles will translate to cyclic metabolic flux. With a resurgence in isotope tracers, in tandem with annotated genome-scale reaction networks, new technologies can be developed to detect transitory metabolism amidst circadian cycles. Improved MS instrumentation can now provide rich isotopologue datasets, however much work remains to process and interpret such complicated data. Specifically, tools must be developed to consider the possible routes through metabolic networks which produce observable isotope enrichments.

The results of these objectives are outlined in the proceeding chapters. The LC-MS metabolomics methods necessary to discern systems-level metabolic cycles are developed in Chapter 2 through a design of experiments approach. The LC-MS methods developed in Chapter 2 are then applied to *Drosophila* in Chapter 3, through a collaboration with Dr. Amita Sehgal at the University of Pennsylvania, to separate metabolite rhythms driven by environmental cues from endogenous clocks. Additionally, comparisons are made to previous metabolomics experiments to gauge conserved metabolite cycles across species. Chapter 4 expands the LC-MS methods developed in Chapter 2 to detect isotope enrichments from a $^{13}C_6$ glucose tracer.

Through continued collaboration with Dr. Sehgal, a new platform is developed to introduce stable isotope tracers to *Drosophila* to uncover novel *in vivo* metabolic flux cycles. Herein are described bespoke computational workflows to process and analyze time-series isotopologue datasets, in addition to dissecting the impact of sleep restriction on rhythms in metabolic flux. As detailed in section 1.4, limitations in sampling resolution and non-steady state conditions in flies preclude quantitative flux analysis, however relative pathway activities can still be ascertained. To address the limited interpretability of isotopologue datasets, both from Chapter 4 and broadly in metabolic flux analysis, new computational tools are developed to model reactions in genome-scale networks at atomic resolution, followed by construction of paths from the glucose tracer to downstream isotopologues of interest, which is described in Chapter 5. This tool is also presented as a new approach to simulate and design isotope tracer studies, with utility beyond chronometabolism.

CHAPTER 2 - Comprehensive optimization of LC-MS metabolomics methods using design of experiments (COLMeD)

Adapted from published work:

Rhoades SD, Weljie AM. Comprehensive optimization of LC-MS metabolomics methods using design of experiments (COLMeD). Metabolomics. 2016;2:183.

2.1. Abstract

Both reverse-phase and HILIC chemistries are deployed for LC-MS metabolomics analyses, however HILIC methods lag behind reverse-phase methods in reproducibility and versatility. Comprehensive metabolomics analysis is additionally complicated by the physiochemical diversity of metabolites and array of tunable analytical parameters. Our aim was to rationally and efficiently design complementary HILIC-based polar metabolomics methods on multiple instruments using Design of Experiments (DoE). We iteratively tuned LC and MS conditions on ion-switching triple quadrupole (QqQ) and quadrupole-time-of-flight (qTOF) mass spectrometers through multiple rounds of a workflow we term COLMeD (Comprehensive optimization of LC-MS metabolomics methods using design of experiments). Multivariate statistical analysis guided our decision process in the method optimizations. LC-MS/MS tuning for the QqQ method on serum metabolites yielded a median response increase of 161.5% ($p<0.0001$) over initial conditions with a 13.3% increase in metabolite coverage. The COLMeD output was benchmarked against two widely used polar metabolomics methods, demonstrating total ion current increases of 105.8% and 57.3%, with median metabolite response increases of 106.1% and 10.3% ($p<0.0001$ and $p<0.05$ respectively). For our optimized qTOF method, 22 solvent systems were compared on a standard mix of physiochemically diverse metabolites, followed by COLMeD optimization, yielding a median 29.8% response increase ($p<0.0001$) over initial conditions. The COLMeD process elucidated response tradeoffs, facilitating improved chromatography and MS response without compromising separation of isobars. COLMeD is

efficient, requiring no more than 20 injections in a given DoE round, and flexible, capable of class-specific optimization as demonstrated through acylcarnitine optimization within the QqQ method.

2.2 Introduction

An ideal metabolomics platform would profile all the metabolites in a living system. Complementary approaches such as GC-MS, LC-MS, and NMR can be employed to enhance analytical coverage of the metabolome, however the high physiochemical diversity in metabolites and technological limitations confine any individual analyses to a relatively small subset of the metabolome. Within LC-MS, multiple column chemistries are also increasingly incorporated into metabolomics workflows to further enhance coverage (Want et al. 2010). Hydrophilic interaction chromatography methods hold significant promise for comprehensive analysis of polar metabolites, however reverse-phase methods are routinely used for small polar molecules (New and Chan, 2008), in part due to a longer history of reproducible chromatography. The choice of MS detection is also critical for augmenting metabolome coverage. For instance, untargeted high-resolution instruments can provide good coverage and sensitivity (Want et al. 2010), while modern triple quadrupole or ion-trap instruments provide ion-switching and fast-scanning capabilities for targeted metabolite identification and quantification (Yuan et al. 2012; Gika et al. 2012; Lv et al. 2011). Holistic improvement of both chromatography and detection parameters requires bespoke methods to address a large multivariate problem space (Gika et al. 2014). This problem has been previously addressed using genetic algorithms and large-scale Bayesian networks (Napoles and Steenbergen, 2014; Correa and Goodacre, 2011), however these approaches do not concurrently optimize numerous parameters inherent in the comprehensive LC-MS methodology, nor have they been applied to HILIC, which is often sensitive to small LC parameter adjustments (Nguyen and Schug, 2008; Hao et al. 2008). An alternative approach for complex method optimization is DoE, which incorporates multivariate modeling of many response variables simultaneously (Eriksson et al. 2006). DoE allows for the manipulation of several factors concurrently and efficiently searches for interaction effects, as opposed to simply changing one factor at a time. Typical DoE workflows start with screening objectives, where the most important factors and their appropriate ranges are chosen and subsequently optimized iteratively. DoE has been used in optimizing other steps of the typical metabolomics workflow, including sample

preparation and data processing (A et al. 2005; Eliasson et al. 2012; Zheng et al. 2013). LC-MS methods have also been improved in this manner, however the response of interest has typically been targeted to one metabolite or a single class of compounds (Zhou et al. 2009; Székely et al. 2012; Kostić et al. 2013; Riter et al. 2005).

Here we demonstrate that a DoE-driven approach has potential for large-scale metabolomics method development by improving metabolome coverage without overtly sacrificing individual metabolite chromatography and MS response. Our main objective was to design and optimize a polar metabolomics platform, while addressing the idiosyncrasies of targeted and untargeted LC-MS metabolomics. We coin this method comprehensive optimization of LC-MS methods through DoE (COLMeD) as a workflow procedure and assess the capability of DoE to improve responses on a diverse set of polar metabolites. We find this workflow is robust to method development across MS detection methods, tailoring the COLMeD approach to an LC-MS ESI+ method using high-resolution qTOF detection subsequent to the initial polarity-switching QqQ optimization. We show that the results are robust to multiple sample types and can be tailored in a class-specific manner by specifically optimizing acylcarnitines from the comprehensive QqQ method. We note improvements over commonly used methods (Yuan et al. 2012; Paglia et al. 2012), and our workflow informed parameter decisions to limit response tradeoffs. Moreover, we describe a generalized procedure, bearing in mind the utility of this approach for efficiently optimizing other facets of analytical method development.

2.3 Methods

Chemicals

All chemical standards used in this study were minimally analytical grade and obtained from commercial sources (Table 2.1). Optima grade acetonitrile and methanol were purchased from Fisher Scientific (Fair Lawn, NJ) for the mobile phase and standard solutions. Optima LC/MS ammonium acetate and formic acid and TraceMetal grade ammonium hydroxide were

used as mobile phase additives and also obtained from Fisher Scientific. All water used in this study was deionized and filtered (18.2MΩ, 0.22µm).

Preparation of standard solutions and biological samples

Standard solutions for positive mode qTOF DoE were prepared as 1mg/mL stocks in 100% methanol and diluted to 1µg/mL in 3:1 acetonitrile:methanol. All samples were centrifuged at 18787g for 5 minutes before LC-MS injection. Gibco horse serum (Invitrogen, Grand Island, NY) and homogenized Drosophila melanogaster samples were prepared using a modified Bligh-dyer extraction (Bligh and Dyer, 1959). Briefly, 120µL of 2:1 methanol:chloroform was added to 20µL of serum or 40mg of fly tissue, followed by a brief vortex and 15 minute sonication. 40µL of both chloroform and water were added to the solution, followed by centrifugation at 18787g for 7 minutes to form the bilayer. The top layer, containing the aqueous fraction, was isolated and dried down overnight. The dried pellet was resuspended in either 100µL or 400µL of 50:50 water:acetonitrile for serum and fly respectively before LC-MS injection.

LC-MS conditions

Chromatographic separations for the ion-switching DoE were performed on an XBridge BEH Amide column (2.1x100mm, 2.5µm, Waters Corporation, Milford, MA) with a 2.1x5mm Vanguard pre-column. DoE chromatography for untargeted qTOF analysis was performed on an ACQUITY UPLC BEH Amide column (2.1x150mm, 1.7µm), with a 0.2µm in-line filter. Both methods utilized an ACQUITY H-Class UPLC (Waters Corporation). The mobile phases for the ion-switching analysis were initially taken from Yuan et al., where the aqueous mobile phase consisted of 95:5 water:acetonitrile with 20mM ammonium acetate and ammonium hydroxide, pH 9, with the organic mobile phase as 100% acetonitrile. Mobile phase A for the qTOF LC-MS method was comprised of 95:5 water:acetonitrile with 2mM ammonium acetate and 0.2% formic acid, while mobile phase B consisted of 90:10 acetonitrile:water with 2mM ammonium acetate and 0.2% formic acid, which was determined through experimental testing as described later in

the text. Mass spectrometry was performed on either a Waters TQD or Waters G2-S qTOF in positive ion mode using ESI+, using Leucine-enkephalin for the lock-mass calibration. As a basis of comparison for our approach, the LC-MS methods described by Yuan et al. (Method 1) and from a Waters HILIC Application Note (Paglia et al. 2012, Method 2) were followed as published. Chromatograms were processed using TargetLynx under MassLynx version 4.1. Statistical analyses and plotting was performed in R version 3.2, and comparisons between DoE rounds were made via paired Wilcoxon signed-rank tests.

Model generation, design, and optimization in the COLMeD process

DoE serves to discover important predictor variables which contribute to one or many desired responses to determine optimal factor tuning (Eriksson et al. 2006). To deal with the complexity of tuning multiple factors to manipulate the many responses in our metabolomics methods, our DoE-driven COLMeD approach employs a partial least squares (PLS) fitting algorithm. Specifically, PLS fits a model to the variation of all responses with the variation of the factors by accounting for their covariance. This method of fitting is more efficient than multiple linear regression (MLR), which is also common in multivariate optimization problems, since MLR fits separate regression models for each response. MLR also suffers when handling missing data points, which we had encountered in our response matrix, given some chromatographic peaks were not always present depending on the LC-MS factor settings. In our case, dependent variables are the original analytical responses (e.g. metabolite peak area and chromatographic values), which are tuned to independent LC-MS factors by transformation into latent variables. The number of latent variables, or PLS components, were determined through the default mechanism in MODDE v11 (Umetrics, Umeå, Sweden), whereby components were added to improve goodness of fit (R^2) until the goodness of prediction (Q^2) was compromised by overfitting the model. The predictive performance of the model was computed via 7-fold cross validation. A major advantage of this approach is the ability to weigh one response more than another across DoE rounds, which we utilize heavily in our COLMeD process. For example, responses with

significant peak area under the curve (AUC) after the first round were downweighted in the predictive modeling to favor LC-MS settings which improve other features with lower responses. AUC responses were set to be maximized in the modeling, while the peak widths were tailored such that peak width objectives were either 4 seconds for the untargeted qTOF method or 15 seconds for the QqQ method, which benefits from slightly wider peaks due to tradeoffs of scanning over hundreds of MRMs. Additionally, the response objectives can be adjusted round over round to achieve iterative improvement. After modeling, a new set of experiments are generated, in the form of LC-MS settings. This process constitutes one round of DoE, which would be repeated until optimal conditions are met (specific COLMeD processes listed in Table 2.2).

In a regularly shaped design region, central composite or full factorial designs are typically chosen to explore the edges of the design space. However, D-Optimal designs generated in MODDE can accommodate experiments with irregular design regions (Eriksson et al. 2006), which allowed us to impose constraints on our LC-MS settings which were not feasible or desirable, for example long LC gradients coupled with high flow rates. We tested the edges of the irregular design space in addition to replicate LC-MS injections in the center of the space to gauge reproducibility and model validity. In addition, we performed a conserved triplicate injection at the end of each DoE round as a quality control measure across batches. To rationalize the LC-MS parameters for the next round of experiments, we utilized both visual representations of optimal regions within the design space and a quantitative optimizer function which generated a list of parameters to yield an optimized solution using the PLS model. In addition to the model statistics and predictive functions of the PLS model, we also evaluated the VIP value, which is a multivariate metric used to identify the relative importance of an original predictor variable (i.e. before transformation) to the model (Eriksson et al. 2006). These values identify non-significant contributions of LC-MS parameters to the metabolite responses, which allowed us to assign fixed values and simplify the design space for the next DoE round. We chose to use MODDE software due to integrated cross-validated model fitting, model fit visualizations, and predictive capabilities.

Open-source platforms for each of the steps in the COLMeD process could be alternatively used to build an in-house workflow.

Ion-Switching COLMeD

While the ultimate goal was to optimize both LC and MS conditions, the initial experimental design space including LC gradients was too large to combine with MS parameters and therefore required a two-stage approach, outlined in Table 2.2. A screening linear objective design of LC-only tuning was chosen for this initial round, omitting interaction effects and requiring only 13 LC-MS injections. Response measurements were taken on multiple sample types, horse serum and homogenized fly samples, which provided an added measure of confidence in designing the next round of experiments. The total initial response optimization consisted of a set of 33 responses in horse serum (Table 2.3), as well as unique responses found in fly tissue but uncommon in serum in order to enhance overall coverage in the initial screen (Table 2.4). Measured response variables were chosen based on criteria designed to elucidate a broad physiochemical range of metabolites and the presence of marginally detectable metabolites, in addition to measures of peak quality. We tailored the responses to reflect particular considerations of MRM-based analysis. For example, glutamine and lysine have overlapping MRMs, thus we fit the PLS model to predict maximal retention time separation. MRM transitions and voltages were optimized by using pure standards or from the METLIN and HMDB mass spectrometry databases (Smith et al. 2005; Wishart et al. 2013). For AUC response optimization, the objective defined in the PLS fitting was set by using the mean AUC from the LC-MS injections of that DoE round as a threshold, from which the optimizer and design space plots were used to find conditions predicted to increase AUC for the subsequent DoE round.

Factors considered for the LC optimization included the initial LC flow rates (0.1-0.25mL/min, continuous variable) and gradient types (1-4, Figure 2.1, discrete). The gradients were rationalized from both published (Yuan et al. 2012) and unpublished work. The LC-MS parameters proposed by Yuan et al. (Gradient 1) served as a starting point to build our

metabolomics platform, however we felt the COLMeD process could improve response within the QqQ method as well as and LC solvents as Yuan et al. for the QqQ method, we relaxed the LC parameters in an exploratory manner before refined tuning in later rounds. The MS parameters were initially set as close to the published settings as possible while aligned with vendor-specific voltage parameters. Initial gradient times (10-16min) were purposefully imposed to achieve reasonably high throughput. The gradient time refers to the time of injection until the wash step.

Modular workflow optimization

After Rounds 2 and 3, we sought to further optimize the method for a specific class of compounds with shared chemical properties, in which case the parent method could be used to tune the response for specific compounds, such as carnitines. The analytical factors were analogous to other rounds of QqQ DoE, but the responses were limited to peak AUCs and peak widths for carnitine, acetylcarnitine, propionylcarnitine, and butrylcarnitine. The data from DoE rounds 2 and 3 were combined and used as inclusions in a D-Optimal quadratic design, whereby only an additional six test runs were needed to finish out the model design. The experimental space in which the method optimum predictions from Rounds 2 and 3 overlapped with predictions based off of these additional test runs with good model statistics confirmed the optimized method and thus completed the class-specific DoE.

Untargeted qTOF COLMeD

Optimizing chromatography for an untargeted method requires additional considerations due to the large number of unknown responses. Rather than optimizing responses on a serum sample, which would contain many unknown features, we initially developed the qTOF method on a standard mix of 48 diverse polar metabolites injected at 1µg/mL (Table 2.1). Prior to DoE, 22 LC solvent systems were compared using the standard mix. These solvents were based on a literature search and are listed in Table 2.5 (Want et al. 2010; Kivilompolo et al. 2013; Ivanisevic et al. 2013; Zhou et al. 2013). Mass spectrometry settings were based on data from the ion-

switching method development, where desolvation temperature and gas flow were set to 500°C and 1000L/Hr respectively. Unless specified by a particular published method, the gradients were aligned across each solvent system, with 45°C column temperature. Given the solvents used in the QqQ method are pH 9, new solvents were required for a qTOF method in positive ionization. Aggregate measures of peak capacity, peak skew, peak resolution between any two pairs of peaks, number of peaks, MS response, and peak widths were all compared to choose the initial LC solvents (Table 2.6). These metrics were used in addition to inspection of chromatography to choose the best solvents manually. We found that some solvents yielded split peaks and prohibited detection of all the metabolites in our mix, thus we felt the need to inspect these results and choose accordingly before proceeding to strictly quantitative optimization of the method using DoE. After selecting the LC solvents, the COLMeD approach was divided into two parts to optimize LC and MS settings separately, which was in large part guided by our QqQ COLMeD findings. Three rounds of DoE were performed for the LC factors (Table 2.7), which included responses for isobar separations of leucine/isoleucine and alanine/sarcosine, with fixed MS parameters. For ease of comparison, peak response (AUC), peak width, and peak skew were converted to rank-based values, whereby each injection was ranked in each of these metrics across every injection from a given DoE round. The holistic peak metrics used as responses in the qTOF COLMeD are more amenable to peak-picking methods used in untargeted data processing algorithms. Our objective with the initial LC DoE was to heavily favor optimization of chromatography, which naturally derives from the variety of peak quality metrics (peak shape, width, separation, etc..) chosen compared to a singular readout of pure metabolite response on the MS. In addition, automated integration of peaks can be difficult when flow rates and gradients are tuned, which does not change during MS factor tuning. AUC was thus one of several responses optimized in the LC DoE, but subsequently the sole response variable used in the MS DoE after chromatography was fixed.

Thus we maintained a similar workflow to the QqQ COLMeD procedure, with additional fit-for-purpose modifications to the developmental process. After LC optimization, two rounds of

DoE were performed for the MS parameters in ESI+. Similar to the polarity switching design, the LC factors consisted of flow rate (0.2-0.5mL/min), gradient slope (4-9, which was calculated by percent change in solvent B divided by gradient time), and column temperature (30-60°C). By having a simpler design space with only three factors, we could employ a more complex a D-Optimal quadratic model (20 runs). We were able to further simplify the design by eliminating non-significant factors and perform rounds 2 and 3 as full factorial designs (12 runs each, in a 2x2x3 design). A full factorial design allows for simultaneously testing three levels of each factor and support a quadratic model. While typically experimentally costly, with only two factors, this design space can be tested with only 12 injections, including center point replicate injections. This design is thus identical to the central composite face-centered (CCF) analysis, which is recommended for full scale investigations and optimization after elimination of less important factors from earlier DoE rounds (Eriksson et al. 2006). Our criteria for removing factors included both analysis of coefficient plots and displaying a VIP score below 1. The detailed models informed tradeoffs in analyte response, while also considering conditions providing sufficient chromatographic resolution between isobars. The MS factors consisted of sampling cone voltage (20-40V), desolvation temperature (400-550°C), source temperature (90-150°C), cone gas flow (20-80L/Hr), and source offset (60-100V). The sole response optimized for the MS DoE was average AUC rank for each injection on the 48 standards. DoE was completed when we were able to identify the LC-MS parameters that met our response thresholds, the elucidation of tradeoffs in the method, and weak PLS model statistics, which indicated a tightly constrained design space with minimal gains for further improvement.

2.4 Results and Discussion

Round 1: Initial LC screening for polarity-switching method

The initial screening batch for the comprehensive quadrupole method LC conditions consisted of 13 injections, repeated for both horse serum and fly samples. Predictive design space plots and optimizer analysis of both data sets yielded similar trends for all three factors

(Figure 2.2). Notably, lower flow rates and/or longer gradients were predicted to improve response, with gradients 3 and 4 yielding predicted design space regions with the most response criteria met. Some differences in the predicted gradient time optimum likely result from slightly different response lists across sample types. However, given both flow rate and gradient time had significant impact (VIPs of 1.43 and 1.08 respectively), also supported by analysis of PLS loadings plot (Figure 2.3), our initial estimates of factor ranges based on a priori rationale and desirability of run time and flow rates needed to be expanded. The model statistics from Round 1 serum analysis gave us confidence in obtaining the method optimum via expanding these factor ranges (Table 2.8).

Continued DoE with LC and MS factors

After adjusting the LC factors from Round 1, the MS factors and column temperature were added to the design. While displaying a relatively weak effect compared to other factors, gradient 4 slightly outperformed gradient 3 (nonsignificantly), and was arbitrarily selected and fixed in future designs. Running multiple optimizations predicted flow rates above 0.15mL/min to improve results. To some degree this prediction contradicts the predictions from Round 1, which may be resolved with increased sampling. Given that flow rate maintained an important contribution to the PLS model (VIP = 1.09), the flow rate was restricted to regions of the design space where the predictive plots and optimization functions overlapped in optimum predictions (0.15-0.3mL/min). Desolvation temperature also had a large effect on response as demonstrated by the largest VIP value (1.19) amongst LC-MS factors. Consequently, this factor was restricted to 300-500°C for Round 3 based on the predictive plots and optimizer function. The predicted optimal gradient times were variable, however given the desire to increase throughput, 12-18 minute gradients were set for Round 3. Column temperature had less effect on the model compared to the other factors (VIP=0.89), and these temperatures were restricted to 40-55°C for Round 3 based on the predictions. Desolvation gas flow was retained as a factor in the subsequent model despite a minimal contribution (VIP=0.8), albeit restricted to maintain

compatibility with the desolvation temperatures according to the manufacturer's recommendations.

Further refinement of design space

The updated factor ranges were used for the fractional factorial design in Round 3, necessitating 17 runs. Predictions from the model yielded slightly shifted but generally consistent results with Round 2. Flow rate and desolvation temperature again were the most significant factors, (VIP of 1.88 and 1.37, respectively). Optimal flow rates gravitated towards the low end of the 0.15-0.3mL/min range. Conversely, higher desolvation temperatures were predicted to perform better. Gradient times were consistent from predictions from Round 2, likewise, column temperature and desolvation gas flow were largely irrelevant, with the factor ranges being unaffected by the updated predictions. Over the course of the three rounds we noticed marked improvement in multiple endpoints, including MS response, peak width (towards our goal of 15sec widths at half height), and number of metabolites (Figure 2.4A).

Acylcarnitine-specific DoE

To explore the notion of targeting subsets of metabolites from the comprehensive parent method, data from rounds 2 and 3 of the polarity switching DoE were combined to generate predictions for increased AUC and optimized peak width of carnitine, acetylcarnitine, propionylcarnitine, and butrylcarnitine. Based on this data, lower flow and longer gradients were expected to improve response and peak width. Analysis of the acylcarnitine-specific LC-MS runs generated from modeling the data in rounds 2 and 3 revealed consistencies in the predictions and improved AUC response and peak width with strong PLS fitting (Table 2.8), demonstrating successful confirmation of the predictions and optimization of a compound class specific method (Figure 2.5).

Refinement and validation of comprehensive method

We primarily attributed improvements in the acylcarnitines across Rounds 2 and 3 to adjustments in desolvation temperature. Given this information, we then evaluated the original 33 responses chosen in the comprehensive method with the acylcarnitine-specific experimental design to observe tradeoffs between response improvement of the carnitines, which exhibit positive ionization, versus other metabolites in both ESI+ and ESI- modes (labeled as Round 4 in Figure 2.4B). While not a true round 4 DoE design for the comprehensive method, we can leverage this extra information to refine our final COLMeD output. We found that between Round 4 and Round 1, which contained the maximal spread of desolvation temperatures, the median carnitine AUC increase was 82.2% compared to 54.9% for all other metabolites. We subsequently adjusted desolvation temperature to 450°C as a final tuning of the comprehensive method which contains both ESI+ and ESI- metabolites. The final parameters for the comprehensive QqQ method were 0.15mL/min flow rate, 20 minute gradient time (Mobile phase B changed from 85-30% over the first 5 minutes and held until the wash step at 20 minutes), 950L/Hr gas flow and 45°C column temperature. The overall COLMeD progression of LC-MS parameters is listed in Table 2.9. Manufacturer's notes from Waters suggest 400°C and 800L/Hr desolvation settings for a 0.15mL/min flow rate, corroborating optimized and safe conditions. Although the flow rate is below the optimum efficiency for a 2.5µm particle size, the COLMeD approach optimized our methods to be fit-for-purpose. Additional injections of horse serum were analyzed at these conditions as a validation measure of the final method (labeled as 'COLMeD Final' in Figure 2.4B). Improvements were noted over Round 1 in both the percent AUC increase (median increase of 161.5%, p=7.76e-16, Figure 2.4B) and the number of metabolites detected, from 163 to 188.

Benchmarking the COLMeD result

We compared our final method to two other well utilized methods: a polarity-switching method utilizing the same solvents and column from Yuan et al., (Method 1) as well as a vendor published method optimized for high-resolution untargeted qTOF analysis compatible with our

instrumentation (Waters) used in several metabolomics studies (Want et al. 2010; Paglia et al. 2012; Bruce et al. 2009, Method 2). We observed a total ion current increase in measured metabolite channels of 105.8% and 57.3% over these two methods respectively, (Figure 2.4C), with a median 106.1% increase in metabolite response over Method 1 (paired comparison, p=1.46e-13) and a median 10.3% increase over Method 2 (p=0.042). We were also able to maintain broad metabolite coverage with the COLMeD output method, yielding 188 metabolites compared to 181 and 190 in Methods 1 and 2 respectively, meeting our overall objective of developing sensitive and deep polar metabolomics methods. The coefficient of variation across replicate injections also decreased for a given metabolite compared to Methods 1 and 2 (p=6.27e-03 and p=7.16e-04 respectively, Figure 2.6), suggesting improved method precision.

Untargeted qTOF DoE

Given the difference between triple quadrupole and untargeted qTOF-based metabolomics, we employed a modified COLMeD approach. Ranked-based metrics (Table 2.6) and visual inspection of chromatography were used to choose the initial LC solvents between 22 solvent combinations in ESI+ mode before further DoE optimization of chromatography (Table 2.5, 2mM ammonium acetate with 0.2% formic acid chosen as the final additives). Of the six interaction variables for LC and MS factors in the PLS model for the Round 4 QqQ DoE, only gradient time with desolvation temperature yielded a VIP score over 1, while four of these six factors had a VIP score below 0.8. Thus we felt the interaction of LC and MS factors were small enough to optimize LC and MS on the qTOF separately (Table 2.10). After the first round of DoE, optimizing LC parameters only, predicted response optimums resided around 0.3-0.4mL/min with 3-5 gradient slope. Both of these factors contributed significantly to the data (VIP values of 1.9 and 1.65), unlike column temperature (VIP=0.62). Consequently, column temperature was fixed at 40°C for Rounds 2 and 3, which facilitated full factorial designs for optimizing flow rate and gradient slope (Figure 2.7). Results were consistent after Round 2, and flow rates were focused to 0.35-0.4mL/min with a gradient slope between 4 and 5. In this process, important tradeoffs

were noted. For example, 0.35mL/min was the minimum flow rate to maintain a Leucine/Isoleucine and Alanine/Sarcosine resolution of at least 1.2. Peak capacity also improved with higher flow rates, but AUC ranks improved towards 0.3mL/min, thus 0.35mL/min was considered a good compromise for these responses (with a 5 gradient slope as defined in the methods and 40°C column temperature chosen as the final conditions, Figure 2.8A). The PLS fitting to data from the Round 1 LC dropped significantly by Round 3, informing us of a highly constrained design space where there existed little room for improvement (Table 2.8). Chromatographic parameters improved with minimal sacrifice to response, particularly in regards to decreasing the spread of peak widths and peak skews (Figure 2.8B). These settings also closely follow the UPLC linear velocity for the 1.7µm column while maintaining safe backpressures. Thus for the final LC conditions, the gradient was changed from 100-20.6% B over 15 minutes at 0.35mL/min, followed by a wash of 100% A for 5 minutes. Mobile phase B was changed from 0-100% from 20-22 minutes and held for column equilibration until 30 minutes. For the MS parameters, analysis of the initial linear screening batch produced a compelling model (Table 2.8) with desolvation temperature as the dominant factor (VIP=2). Positive correlation of AUC and desolvation temperature is also in line with the QqQ method optimization. Desolvation temperature was then fixed for Round 2 to reduce design complexity. Source offset was the only other factor which weighed significantly in the PLS models (VIP=1.75), improving AUC rank at minimum voltages. The other factors had low VIP scores (all below 0.9) and nonsignificant coefficient values, though the trends in improving response were consistent in both rounds. Two rounds of DoE were considered sufficient to improve response to complete the LC-MS optimization of the untargeted method in a highly efficient manner. Round 2 yielded an average response increase of 29.8% (p=3.016e-05), while the chosen parameters within the Round 2 factor settings only yielded a 2.9% increase over the average AUC for the entire round, indicating minimal room for further improvement. Final conditions were 550°C desolvation temperature, 25V cone voltage, 60V source offset, 120°C source temperature, and 50L/Hr cone gas flow (Figure 2.9).

2.5 Concluding Remarks

In this study, the COLMeD approach is demonstrated as an efficient and flexible tool to optimize multiple LC-MS metabolomics methods with different objectives. This may be particularly useful due to differences between results an individual lab compared to literature parameters. Improvements from the starting points of the LC-MS design space were noted for metabolite responses and their chromatography using a limited number of injections. We also noted improvements in these responses over other established polar metabolomics methods. We do acknowledge that while all columns tested in this study and the benchmarked methods have an amide chemistry, the column dimensions vary which may impact the data. A subset of metabolites with similar chemical properties were further optimized within the comprehensive method, which could allow for acylcarnitine-specific analyses without needing to switch solvents or columns. We feel this modular approach can efficiently optimize analyses of other particular metabolite groups of interest. Tailoring the COLMeD approach for untargeted metabolomics on the qTOF also yielded improved chromatography and response while maintaining sufficient isobar separation. Consequently, it is important to note that there are many ways to optimize these methods within the COLMeD framework. One could optimize only LC factors, using more thorough designs after the screening round to add confidence in obtaining the method optimum, followed by a similar workflow for the MS factors, which was the route was taken for the untargeted qTOF method development. This approach may require more injections and time, but has the advantage of being amenable to more automated methods for peak analysis once the chromatography is fixed. Conversely, we felt our combined LC-MS linear model approach for the polarity-switching method could yield our desired output while minimizing injections, given most of the coefficients for the interaction variables to be among the lowest in both rounds 3 and 4 of the QqQ COLMeD process (Table 2.10). A combination of more detailed model designs, along with a smaller initial design space, will likely yield a stronger predictive model, as we had found with the separated LC and MS optimizations (Table 2.8). We generally recommend separating LC and MS optimizations if time allows and additional rigor is required, however without parallel analysis of

combined versus sequential LC-MS optimizations, we cannot definitively say if the response gains would be significant. In addition, broader assessments of peak quality and response are more suitable for optimizing untargeted metabolomics methods. In our case, optimizing these aggregated metrics met our objective of developing comprehensive metabolomics methods, as opposed to more focused and quantitative metrics such as limits of detection and response variance. Others have recently used a Derringer function approach to comprehensively assess peak characteristics in a multi-analyte mixture (Sampsonidis et al. 2015), similar to what we employed here. We stress that during optimization of complex methods over many responses, not every response can be maximally improved and thus requires analysis of tradeoffs. However, one benefit in utilizing these statistical models is the ability to stress the optimization of more important metabolites or responses. We also note that in method development for LC-MS metabolomics, the method optimum is largely defined by the user, and thus the response selection must be fit for the experimental purpose. Future studies will further validate these methods with different sample types and responses, such as the compounds found in our homogenized fly samples but not horse serum. More importantly, we have laid out a thorough description of the COLMeD workflow from which we hope can be useful in not only LC-MS metabolomics but other complex method types which require adjusting multiple factors to optimize multiple responses.

Acknowledgements

The authors would also like to thank Saikumari Krishnaiah for assistance with qTOF data acquisition and Barry Slaff for fruitful discussions regarding qTOF data analysis. S.D.R. is supported through a Pharmacology T32 Training Grant (T32 GM008076). Supported in part by the Institute for Translational Medicine and Therapeutics (ITMAT) Transdisciplinary Program in Translational Medicine and Therapeutics.

Table 2.1 Metabolite standards used in qTOF method development, with associated HDMB and KEGG identifiers

Metabolite	Human Metabolome Database	KEGG	Source
Glycine	HMDB00123	C00037	Fisher Scientific
Taurine	HMDB00251	C00245	Fisher Scientific
Sucrose	HMDB00258	C00089	Fisher Scientific
Alanine	HMDB00161	C00041	Sigma
Serine	HMDB00187	C00065	Sigma
Proline	HMDB00162	C16435	Sigma
Valine	HMDB00883	C00183	Sigma
Threonine	HMDB00167	C00188	Sigma
Aspartate	HMDB00191	C00049	Sigma
Lysine	HMDB00182	C00047	Sigma
Glutamine	HMDB00641	C00064	Sigma
Glutamate	HMDB00148	C00025	Sigma
Histidine	HMDB00177	C00135	Sigma
Phenylalanine	HMDB00159	C02057	Sigma
Citrulline	HMDB00904	C00327	Sigma
Tyrosine	HMDB00158	C00082	Sigma
Guanosine	HMDB00133	C00387	Sigma
Glutathione-reduced	HMDB00125	C00051	Sigma
Adenosine monophosphate	HMDB00045	C00020	Sigma
Adenosine disphosphate	HMDB01341	C00008	Sigma
Adenosine triphosphate	HMDB00538	C00002	Sigma
Folic acid	HMDB00121	C00504	Sigma
Choline	HMDB00097	C00114	Acros Organics
Betaine	HMDB00043	C00719	Acros Organics
Cysteine	HMDB00574	C00097	Acros Organics
Leucine	HMDB00687	C00123	Acros Organics
Isoleucine	HMDB00172	C00407	Acros Organics
Asparagine	HMDB00168	C00152	Acros Organics
Acetylcholine	HMDB00895	C01996	Acros Organics
Methionine	HMDB00696	C01733	Acros Organics
Hypoxanthine	HMDB00157	C00262	Acros Organics
Carnitine	HMDB00062	C00318	Acros Organics
Arginine	HMDB00517	C00062	Acros Organics
Glucose	HMDB00122	C00031	Acros Organics
Fructose	HMDB00660	C10906	Acros Organics
Acetylcarnitine	HMDB00201	C02571	Acros Organics
Tryptophan	HMDB00929	C00078	Acros Organics
Pantothenate	HMDB00210	C00864	Acros Organics

cyclic AMP	HMDB00058	C00575	Acros Organics
Cortisol	HMDB00063	C00735	Acros Organics
NAD+	HMDB00902	C00003	Acros Organics
Sarcosine	HMDB00271	C00213	Acros Organics
Phosphocreatine	HMDB01511	C02305	Acros Organics
Creatine	HMDB00064	C00300	Acros Organics
Malonylcarnitine	HMDB02095	Not Found	Cambridge Isotopes
Hydroxyisovalerylcarnitine	HMDB13132	Not Found	Cambridge Isotopes
Acetyl-CoA	HMDB01206	C00024	Cambridge Isotopes
Octanoylcarnitine	HMDB00791	C02838	Tocris Biosciences
Palmitoylcarnitine	HMDB00222	C02990	Tocris Biosciences

Table 2.2 COLMeD Workflow for both QqQ and qTOF Methods

Workflow	QqQ	qTOF
Initial factor range and response selection	Factors: LC only (round 1), followed by LC-MS Responses: Metabolites from horse serum (Table S2)	Factors: LC and MS optimization separate Responses: Standard mix, both LC and MS response (Tables S5 and S6)
Pilot LC-MS batch	D-Optimal linear screening design: 13 runs	D-Optimal quadratic design for LC: 20 runs D-Optimal linear for MS: 13 runs
Data processing and fitting to PLS models	Main effect plots and PLS loadings for data assessment	
Model analysis and optimum predictions	Predictive design space plots, and optimizer function for optimum predictions, VIP > 1 for important factors	
Update factor settings and constrain design space for next round; change design as needed	1 round for LC (13 runs) 3 rounds for LC-MS (13, 17, 9 runs)	3 rounds for LC (20, 12, 12 runs) 2 rounds for MS (13, 12 runs)

Table 2.3 List of Responses Optimized in Horse Serum

AUC Response[a] (maximize)	**Peak Width[b] (15sec)**	**Peak Presence[c] (maximize)**
Acetoacetate	Acetylcarnitine	Citrate
Acetylcarnitine	Arginine	Fumarate
Aconitate	Butrylcarnitine	
Butrylcarnitine	Carnitine	
Carnitine	Choline	**Peak Quality[d] (maximize)**
Choline	Disaccharide	
Citrulline	Glycerophosphocholine	Histidine
Disaccharide	Histidine	Succinate
Glucose	Lactate	
Glycine	Ornithine	
Hypoxanthine	Propionylcarnitine	**RT Difference[e] (maximize)**
Malate	Serine	
Ornithine	Tyrosine	Glutamine/Lysine
Propionylcarnitine		
Urea		

[a]AUC represents area under the curve of integrated peak. [b]Width defined by full width half height of the peak. [c]Response given 0 or 1 based on peak presence. [d]Peak quality defined qualitatively, whereby 0 is nonexistant peak, 1 is noisy signal, 2 is defined peak, 3 is Gaussian peak. [e]RT = Retention Time. Brackets in headings indicate the objective criteria used in the optimization functions

Table 2.4 List of Responses Optimized in Fly

AUC Response[a]	**Peak Width**[b]
Acetoacetate	Acetylcarnitine
Acetylcarnitine	Arginine
Aconitate	Butrylcarnitine
ADP	Carnitine
ATP	Choline
Butrylcarnitine	Disaccharide
Carnitine	Glycerophosphocholine
Choline	Histidine
Disaccharide	Lactate
Fumarate	Ornithine
Glucose	Propionylcarnitine
Glycine	Serine
Glutathione-Reduced	Tyrosine
Hexose-6-Phosphate	
Histidine	**Peak Quality**[c]
Malate	Hypoxanthine
Ornithine	
Propionylcarnitine	**RT Difference**[d]
Ribulose-5-Phospahte	Glutamine/Lysine
Succinate	

[a]AUC represents area under the curve of integrated peak. [b]Width defined by full width half height of the peak [c]Peak quality defined qualitatively, whereby 0 is nonexistant peak, 1 is noisy signal, 2 is defined peak, 3 is Gaussian peak. [d]RT = Retention Time

Table 2.5 Twenty-two solvent combinations used in the qTOF LC method development

Method	Aqueous Solvent	Organic Solvent
1	100% water, 0.1% formic acid	100% acetonitrile, 0.1% formic acid
2	100% water, 10mM ammonium acetate, 0.2% formic acid	95:5 acetonitrile:water, 10mM ammonium acetate, 0.2% formic acid
3	100% water, 5mM ammonium formate, 5mM ammonium acetate, 0.2% formic acid	95:5 acetonitrile:water, 5mM ammonium formate, 5mM ammonium acetate, 0.2% formic acid
4	100% water, 10mM ammonium formate, 0.1% formic acid	90:10 acetonitrile:water, 10mM ammonium formate, 0.1% formic acid
5	100% water, 10mM ammonium formate, 0.2% formic acid	90:10 acetonitrile:water, 10mM ammonium formate, 0.2% formic acid
6	100% water, 10mM ammonium formate, 10mM ammonium acetate, 0.2% formic acid	90:10 acetonitrile:water, 10mM ammonium formate, 10mM ammonium acetate, 0.2% formic acid
7	100% water, 0.2% formic acid	100% acetonitrile, 0.2% formic acid
8	100% water, 5mM ammonium acetate, 0.2% formic acid	90:10 acetonitrile:water, 5mM ammonium acetate, 0.2% formic acid
9	100% water, 5mM ammonium formate, 5mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 5mM ammonium formate, 5mM ammonium acetate, 0.1% formic acid
10	100% water, 10mM ammonium formate, 10mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 10mM ammonium formate, 10mM ammonium acetate, 0.1% formic acid
11	100% water, 10mM Ammonium Acetate, 0.2% formic acid	90:10 acetonitrile:water, 10mM ammonium acetate, 0.2% formic acid
12	100% water, 10mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 10mM ammonium acetate, 0.1% formic acid
13	100% water, 5mM ammonium formate, 0.2% formic acid	90:10 acetonitrile:water, 5mM ammonium formate, 0.2% formic acid
14	100% water, 1mM ammonium acetate, 0.2% formic acid	90:10 acetonitrile:water, 1mM ammonium acetate, 0.2% formic acid
15	100% water, 2mM ammonium formate, 0.1% formic acid	90:10 acetonitrile:water, 2mM ammonium formate, 0.1% formic acid
16	100% water, 2mM ammonium formate, 2mM ammonium acetate, 0.2% formic acid	90:10 acetonitrile:water, 2mM ammonium formate, 2mM ammonium acetate, 0.2% formic acid
17	100% water, 2mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 2mM ammonium acetate, 0.1% formic acid
18	100% water, 1mM ammonium formate, 0.2% formic acid	90:10 acetonitrile:water, 1mM ammonium formate, 0.2% formic acid
19	100% water, 1mM ammonium formate, 1mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 1mM ammonium formate, 1mM ammonium acetate, 0.1% formic acid
20	100% water, 2mM ammonium formate, 2mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 2mM ammonium formate, 2mM ammonium acetate, 0.1% formic acid
21	100% water, 5mM ammonium acetate, 0.1% formic acid	90:10 acetonitrile:water, 5mM ammonium acetate, 0.1% formic acid

22	100% water, 2mM ammonium acetate, 0.2% formic acid	90:10 acetonitrile:water, 2mM ammonium acetate, 0.2% formic acid

Table 2.6 Metrics used to choose the optimal starting LC solvent composition among 22 solvent possibilities for untargeted qTOF LC-MS optimization prior to DoE

Metric	Description	Objective
Percentage of peaks under retention 2	The percentage of total peaks detected that had a retention index under 2, calculated using $k=t_R-t_0/t_0$, where tR is the peak retention time and t_0 is the void volume	Minimize
Percentage of peaks under resolution 1.5	Using a combinations function, all possible peak resolutions were calculated between two peaks, defined by $Rs=2(t_{R2}-t_{R1})/(w_{b2}+w_{b1})$, where $_{R2}$ and $_{R1}$ are retention times of two peaks, and $_{b2}$ and $_{b1}$ are base peak widths of two peaks	Minimize
Average peak resolution	Average of all possible peak resolutions between two peaks	Maximize
Average peak skew	Average skew of all peaks, calculated using MassLynx	Minimize
Average peak kurtosis	Average kurtosis of all peaks, calculated using MassLynx	Maximize
Peak capacity using base width	Number of peaks theoretically to separate in a method, defined by $Pc=1+(t_g/w_b)$, where t_g is gradient time and w_b is peak width	Maximize
Average AUC	Average AUC for all peaks	Maximize
Number of peaks	Total number of peaks detected for the standard mix of 48 metabolites	Maximize
Average FWHH	Average peak width at half height	Target at 4 seconds

Table 2.7 Metrics used in the LC qTOF DoE optimization. The objective refers to the objective criteria used in the optimizer functions

Metric	Description	Objective
AUC Rank	Ranks were given to each injection in the DoE round for every metabolite's response, then averaged to give a composite AUC rank for each injection	Maximize
Average FWHH	Peak width at half height averaged	Target at 4 seconds
Average peak skew rank	Peak skew ranks given to each injection in the DoE round for every metabolite's peak, then averaged to give a composite skew rank for each injection (skew calculated using MassLynx)	Maximize
Average peak width rank	Peak width at half height ranks given to each injection in the DoE round for every metabolite's peak by finding the difference from the 4 second width objective, then averaged to give a composite width rank for each injection	Maximize
Leucine/Isoleucine Resolution	Resolution between leucine and isoleucine peaks	Target at 1.5
Alanine/Sarcosine Resolution	Resolution between alanine and sarcosine peaks	Target at 1.5
Percentage of peaks under Resolution 1.5	Using a combinations function, all possible peak resolutions were calculated between two peaks, defined by $Rs=2(t_{R2}-t_{R1})/(w_{b2}+w_{b1})$, where R2 and R1 are retention times of two peaks, and b2 and b1 are base peak widths of two peaks	Minimize
Percentage of peaks under retention 2	The percentage of total peaks detected that had a retention index under, calculated using $k=t_R-t_0/t_0$, where tR is the peak retention time and t0 is the void volume	Minimize
Peak capacity using base width	Number of peaks theoretically to separate in a method, defined by $Pc=1+(t_g/w_b)$, where t_g is gradient time and w_b is peak width	Maximize

Table 2.8 Model parameters across both QqQ and qTOF COLMeD processes. The number of inputs corresponds to the predictor variables, while the outputs are the response variables to be optimized using COLMeD. PLS components indicates the number of components fit to the data for that DoE round. R^2 is the variance explained for the predictor variables in the model, while Q^2 tests the goodness of fit

DoE Round	Inputs	Responses	LC and/or MS Parameters	PLS Components	R^2	Q^2
Round 1 Horse Serum (QqQ)	37	34	3	5	0.727	0.35
Round 1 Fly (QqQ)	38	35	3	3	0.583	0.242
Round 2 Horse Serum (QqQ)	39	33	6	6	0.722	0.112
Round 3 Horse Serum (QqQ)	38	33	5	10	0.781	0.163
Carnitines Horse Serum (QqQ)	13	8	5	7	0.812	0.381
Round 4 Horse Serum (QqQ)	38	33	5	9	0.671	0.293
LC Round 1 (qToF)	12	8	3	7	0.946	0.658
LC Round 2 (qToF)	11	8	2	5	0.857	0.312
LC Round 3 (qToF)	11	8	2	2	0.592	0.185
MS Round 1 (qToF)	6	1	5	2	0.932	0.709
MS Round 2 (qToF)	5	1	4	2	0.68	0.298

Table 2.9 COLMeD factor settings for comprehensive ion-switching analysis.

Factor	Round 1	Round 2	Round 3	Final
Flow rate (mL/min)	0.1-0.25	0.05-0.3	0.15-0.3	0.15
Gradient	1-4	3-4	4	4
Time (min)	10-16	12-20	12-18	20
Desolvation temp (°C)	Not tested	200-650	300-500	450
Desolvation gas (L/Hr)	Not tested	750-1100	800-1100	950
Column temp (°C)	Not tested	30-60	40-55	45

Table 2.10 VIP scores for all factors after Round 4 of the QqQ COLMeD process, including interaction variables.

Factor	VIP Score
Flow Rate (LC)	1.61
Desolvation Temperature (MS)	1.43
Gradient Time (LC)	1.37
Gradient Time*Desolvation Temperature (LC*MS)	1.23
Flow Rate*Gradient Time (LC*LC)	1.14
Flow Rate*Desolvation Temperature (LC*MS)	0.97
Desolvation Temperature*Desolvation Flow (MS*MS)	0.86
Column Temperature (LC)	0.83
Gradient Time*Desolvation Flow (LC*MS)	0.79
Flow Rate*Desolvation Flow (LC*MS)	0.74
Flow Rate*Column Temperature (LC*LC)	0.70
Desolvation Flow (MS)	0.69
Desolvation Flow*Column Temperature (MS*LC)	0.64
Desolvation Temperature*Column Temperature (MS*LC)	0.63
Gradient Time*Column Temperature (LC*LC)	0.61

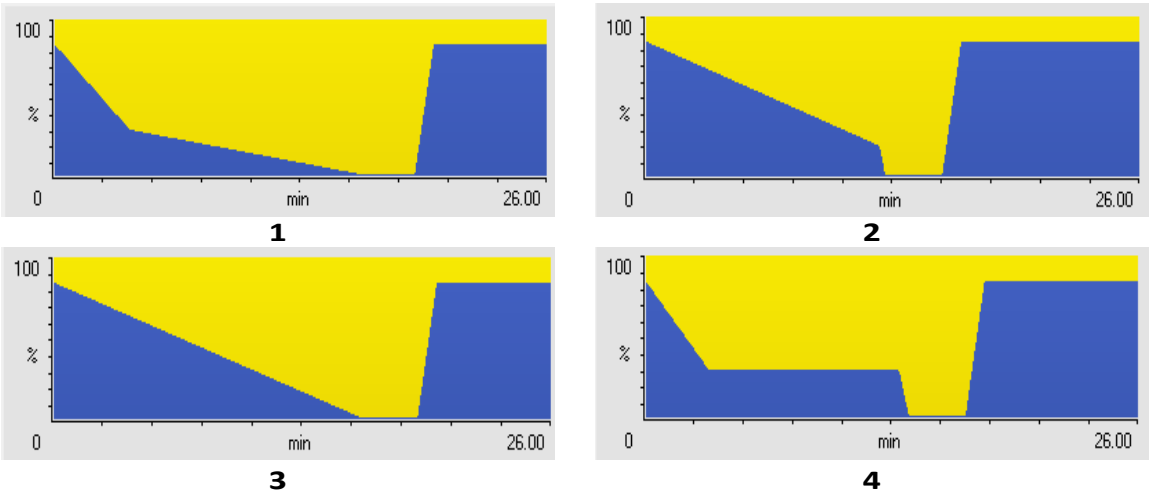
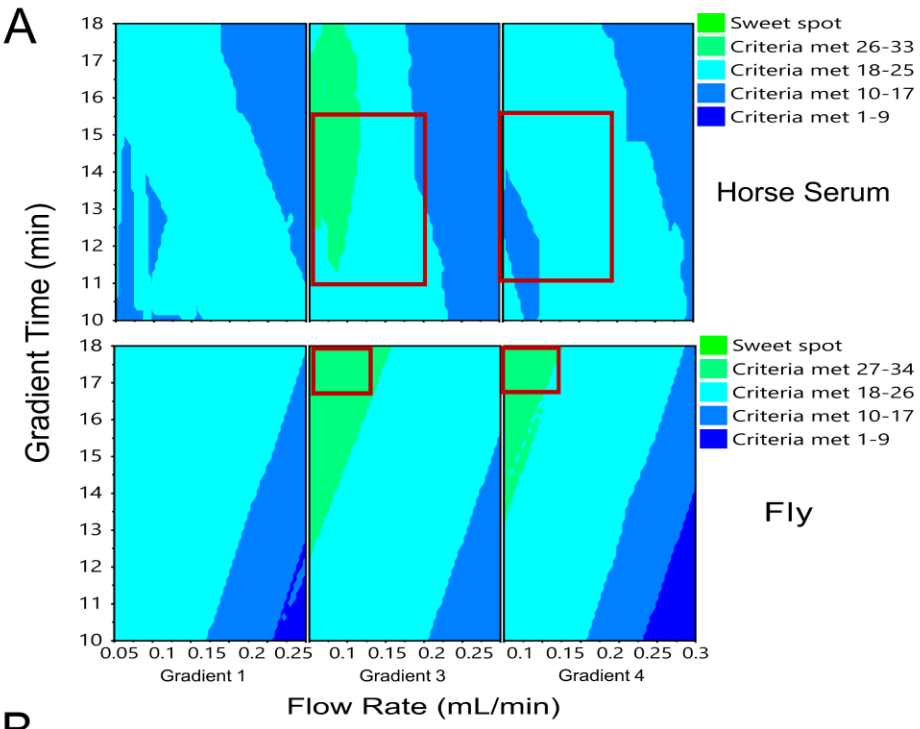


Figure 2.1 Depiction of initial LC gradients, blue indicates organic solvent, yellow indicates aqueous.

A



B

Factor	Initial Settings	Updated Ranges for Horse Serum Using Optimizer	Updated Ranges for Fly Samples Using Optimizer
Flow Rate	0.1 - 0.25	0.05 – 0.2	0.05 – 0.1
Gradient Type	1-4	3 or 4	3 or 4
Gradient Time	10 - 16	11 - 15	17 - 20

Figure 2.2 (A) Predictive plots displaying design space regions with predicted optimal response, based on initial LC screening (round 1) with horse serum and fly samples. Green indicates design space where a maximal number of endpoint response thresholds are predicted to be met. Red boxes indicate regions predicted to improve responses for the next round. (B) Factor settings before and after analysis of serum and fly samples using the optimizer function. These complementary approaches rationalize LC–MS parameters for the next round of experiments.

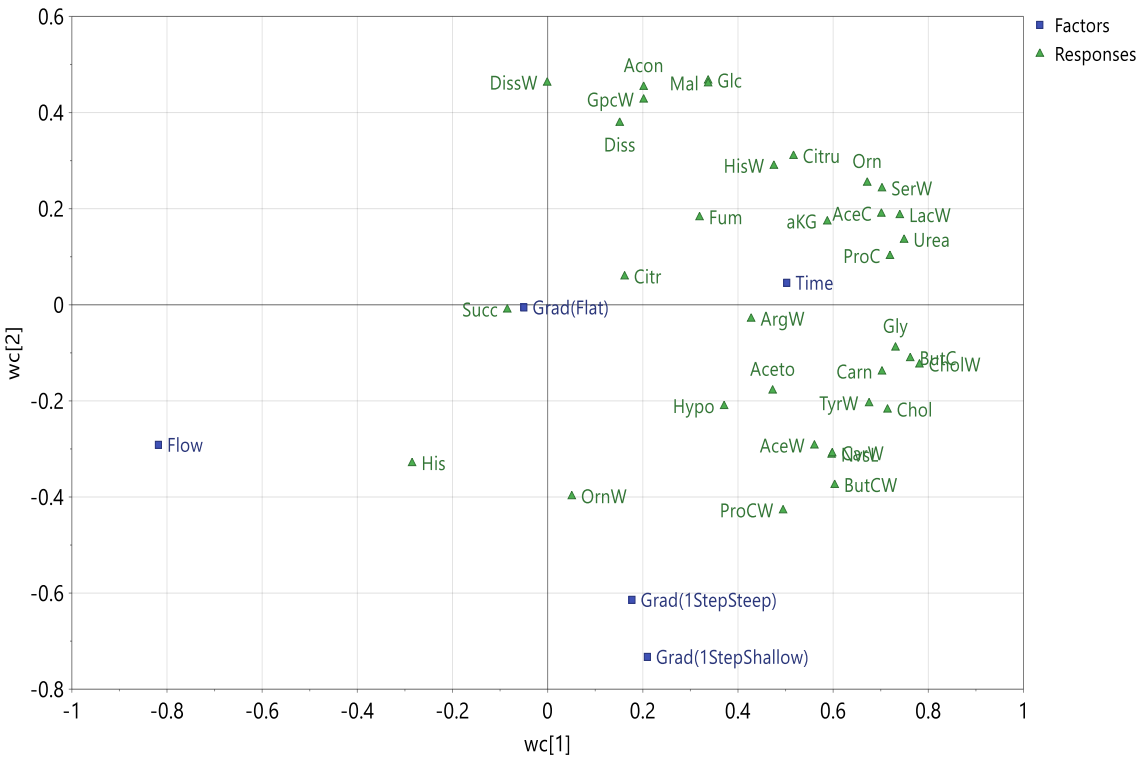


Figure 2.3 PLS Loadings plot after Round 1 of the ion-switching DoE, demonstrating the predicted response improvement by decreasing flow rate and increasing gradient time for subsequent rounds. In reference to Figure 2.1, 1StepShallow is gradient 2, 1StepSteep is gradient 3, and the Flat gradient is number 4.

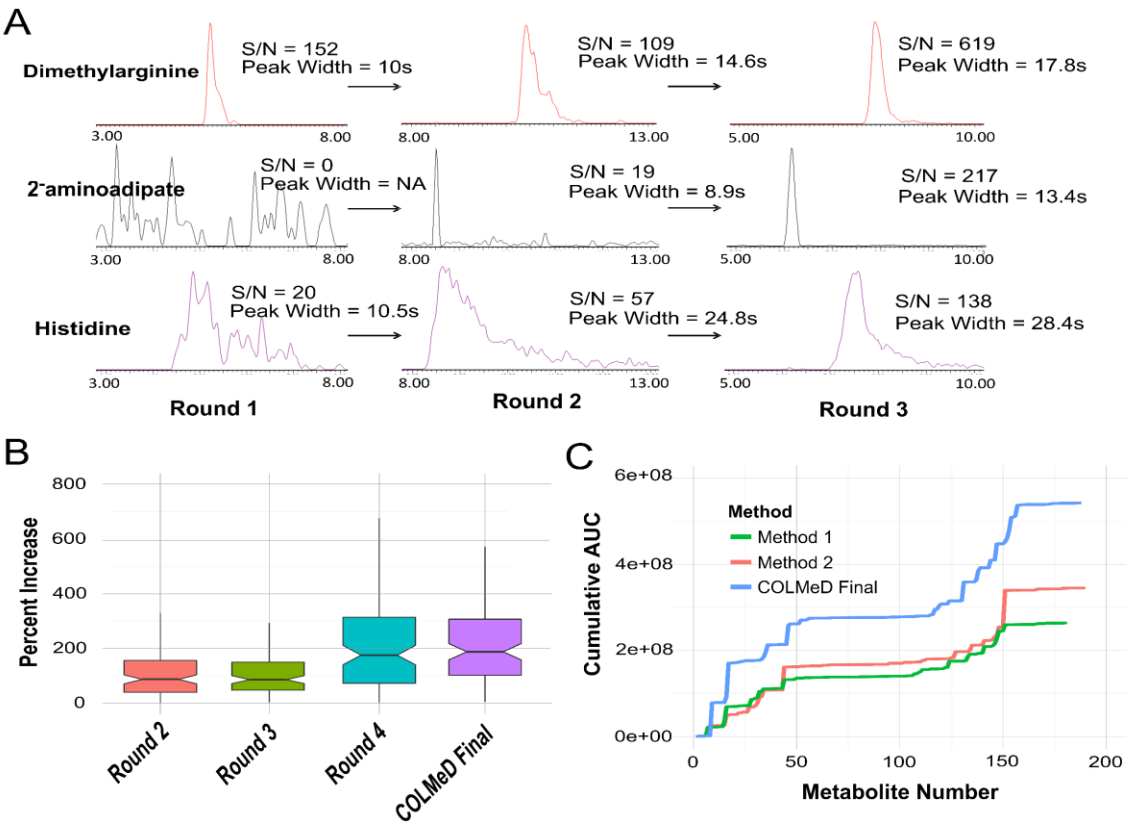
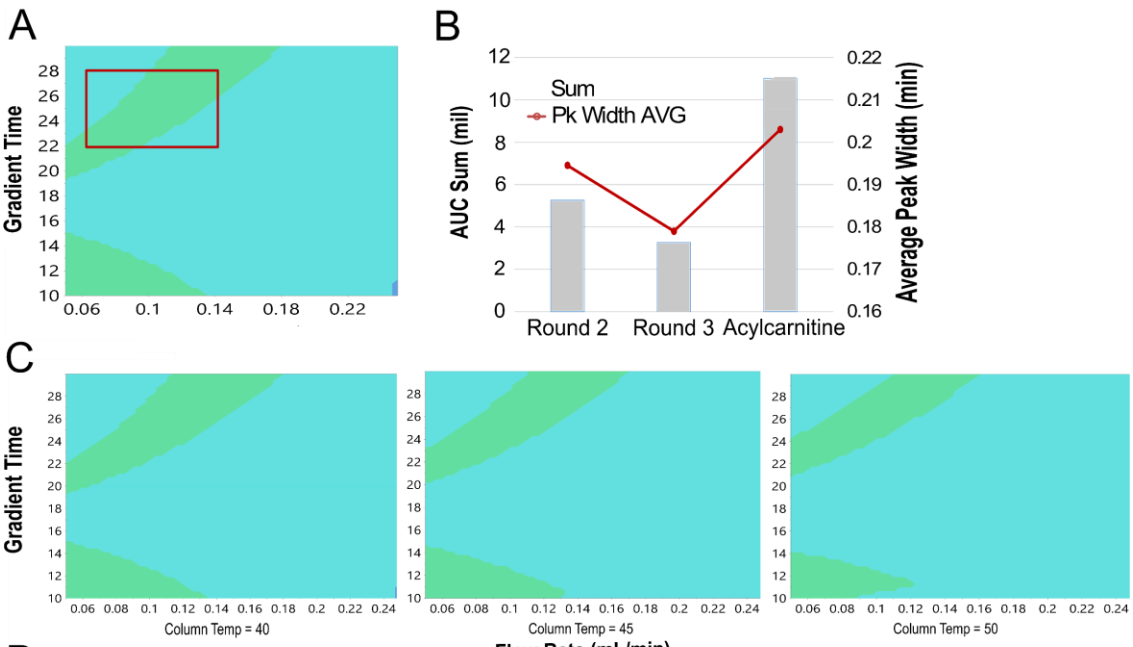


Figure 2.4 (A) Chromatographic improvement in selected metabolites after three rounds of DoE in the QqQ COLMeD process. Signal-to-noise (S/N) and peak width (full width half height) measurements were generated from vendor software after limited peak smoothing and integration. (B) Percent response increase by metabolite across each round. (C) Cumulative AUC plots for QqQ COLMeD benchmarking.



Factor	Initial Settings	Updated Ranges Before Acylcarnitine-specific Analysis	Updated Ranges After Acylcarnitine-specific Analysis	Final Conditions
Flow Rate	0.15 – 0.3	0.1 – 0.18	0.05 – 0.15	0.1
Gradient Type	4	4	4	4
Gradient Time	12 – 18	21-26	24 – 30	25
Desolvation Temperature	300 to 500	600 – 650	450 – 550	500
Desolvation Gas Flow	800 to 1100	900 – 1100	800 - 1000	900
Column Temperature	40 - 55	40 - 60	40 - 45	40

Figure 2.5 (A) Predictive design space plot with optimal response regions for short-chain acylcarnitines, based on combined results of rounds 2, 3, and acylcarnitine-specific batch. Green indicates predicted optimal conditions, while the red box represents overlap of optimal prediction based on only data from rounds 2 and 3, suggesting a well-validated design space. (B) AUC improvement (in millions of counts) in the acylcarnitine-specific DoE round, in addition to peak width nearing the objective set at 0.25min. (C) Design space plots depicting regions of improved acylcarnitine response at 550°C desolvation temperature, and across multiple column temperatures. Column temperature did not have a significant effect on the model, as shown when varied from 40-50°C. (D) Factor ranges predicted to yield optimal acylcarnitine response before and after class-specific DoE. The final conditions were overlaps between predictive design space plots and optimizer function results.

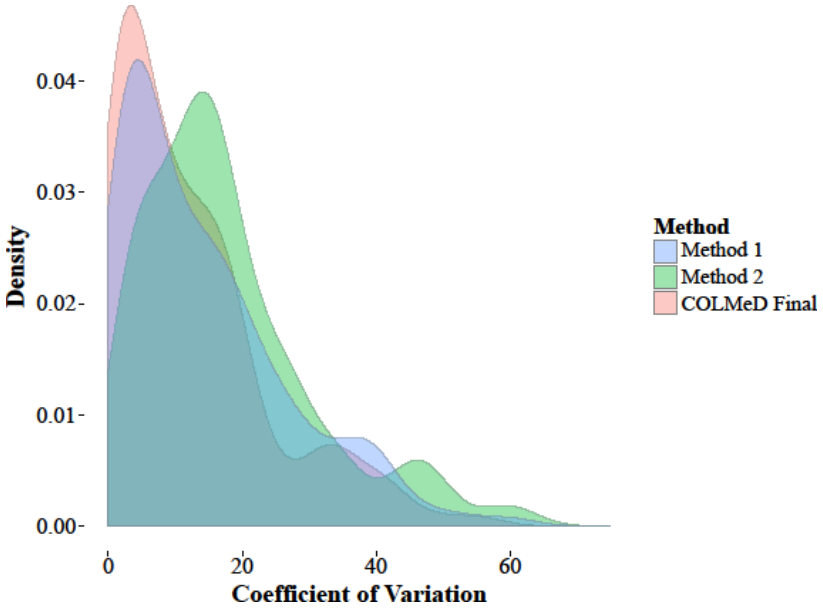


Figure 2.6 Density plot comparing the coefficient of variation for a given metabolite across replicate injections.

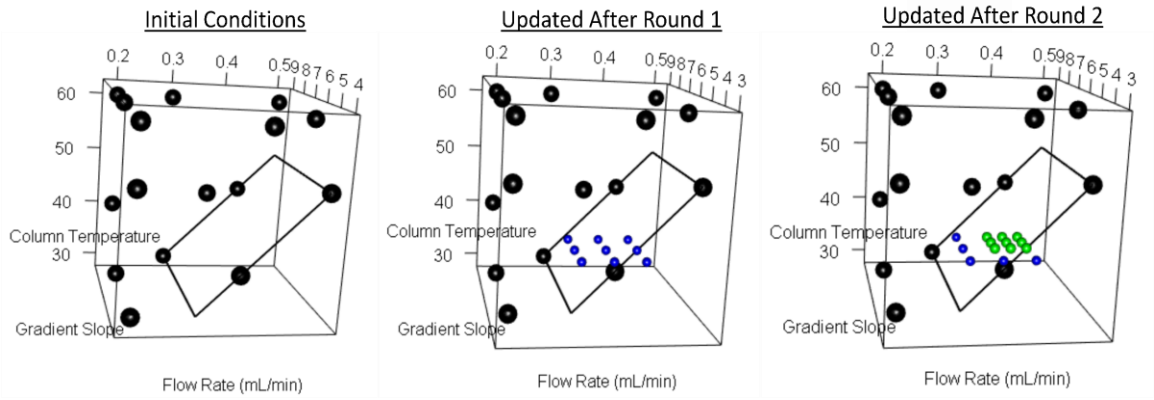


Figure 2.7 Visualization of the design space changes across each LC qTOF DoE round, including the reduction of 3-D to 2-D design spaces formed after fixing column temperature. The black plane within the box depicts an initial constraint on the space. The black, blue, and green spheres represent the parameter settings tested for each round. The final conditions were chosen within the plane of the green dots.

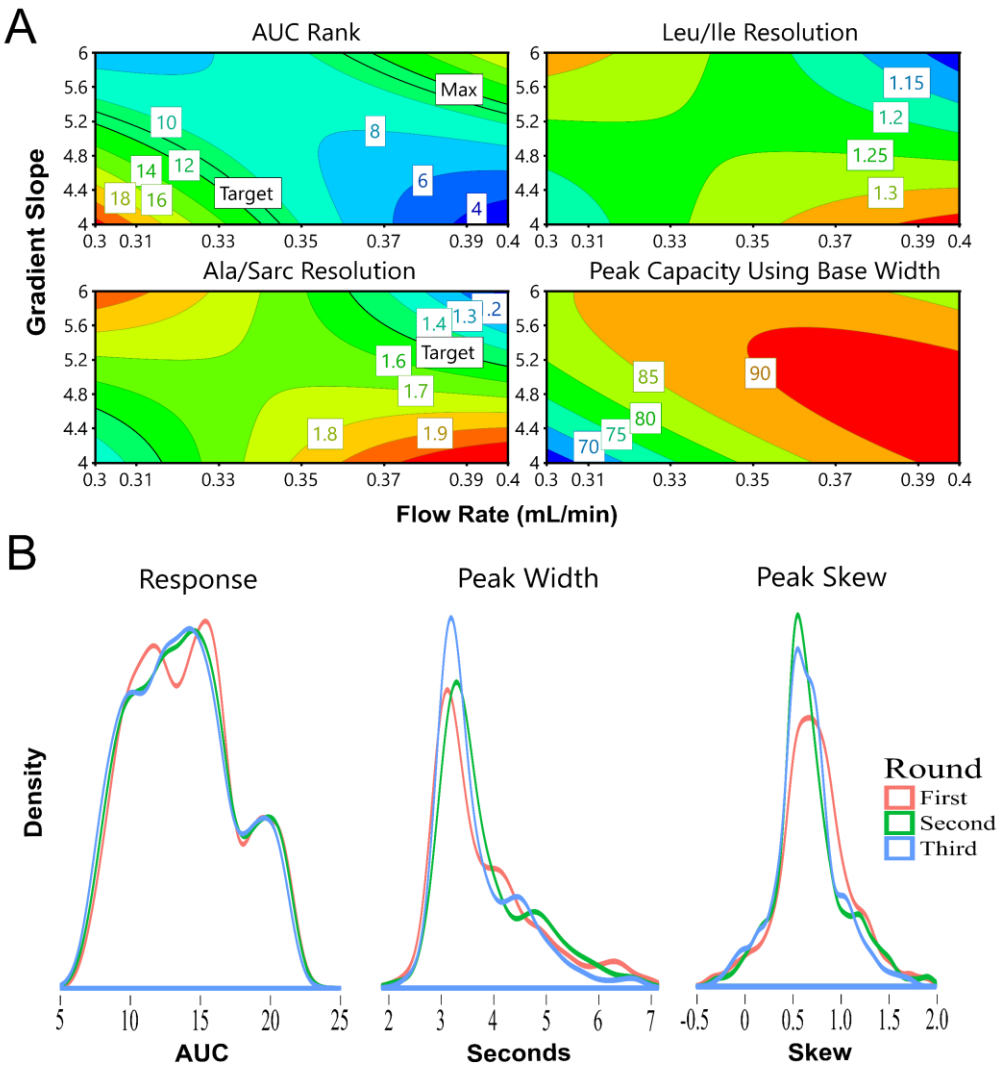


Figure 2.8 (A) Response contour plots after round 3 of LC DoE for untargeted qTOF metabolomics, demonstrating the tradeoffs to be considered in finding a method optimum. (B) Decreased spread of peak widths and peak skews (with a target objective of 4sec widths and minimized skews) across three rounds. These improvements are without compromising the MS response (depicted here as log-transformed AUC values).

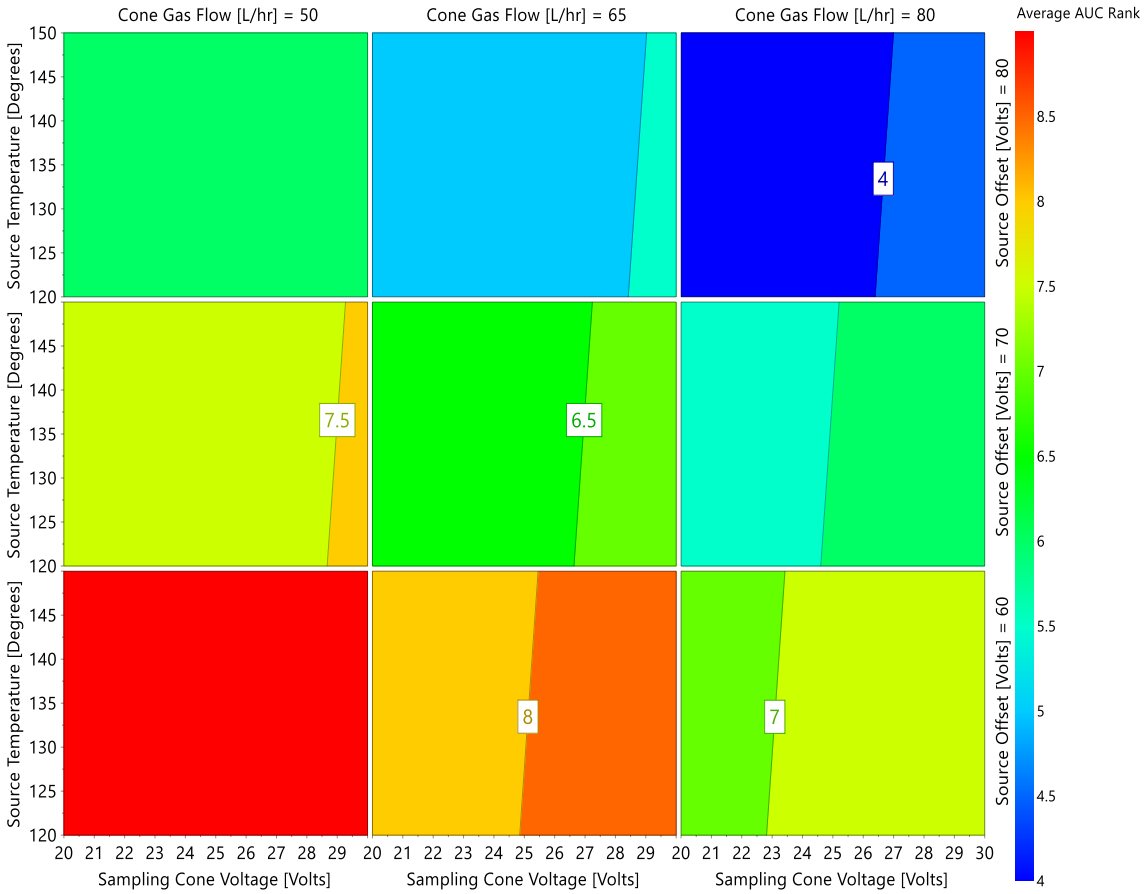


Figure 2.9 4D response contour plot demonstrating design space regions of optimal average AUC rank for qTOF MS DoE.

CHAPTER 3 – Circadian- and light-driven metabolic rhythms in Drosophila melanogaster

Submitted As:

Rhoades, SD, Nayak, K, Zhang, S, Sehgal, A, Weljie, AM. Circadian and light-dependent metabolic rhythms in *Drosophila melanogaster* using high-resolution metabolomics.

3.1 Abstract

Complex interactions of environmental cues and transcriptional clocks drive rhythmicity in organismal physiology. Light directly affects the circadian clock, however little is known about its relative role in controlling metabolic variations *in vivo*. Here we employ high time-resolution sampling in *Drosophila* to uncover over 14% of detected metabolites with circadian periodicity under light:dark (LD) cycles. Many metabolites peak shortly after lights-on, suggesting responsiveness to feeding rather than anticipation as seen in transcriptomics analyses. Roughly 9% of metabolites uniquely oscillate under constant darkness (DD), suggesting clock-driven metabolite rhythms. Strikingly, metabolome differences between LD and DD were only observed during the light phase, highlighting the importance of photic input. Clock mutant flies exhibited only strong ultradian rhythms, including four carbohydrates with circadian periods in wild-type flies. Additionally, conserved rhythms in amino acids, keto-acids, and sugars across species provide a basis for exploring the chronometabolic connection with powerful genetic tools in *Drosophila*.

3.2 Introduction

Diurnal patterns in organismal physiology arise from a complex interaction of environmental inputs, such as light and food, with endogenous circadian clocks. Synchrony of these clocks with predictable environmental patterns confers a selective advantage, particularly to maintain metabolic homeostasis amidst anticipatory activity and feeding behaviors (Edery, 2000). The intersection of metabolism and circadian rhythms is of increasing interest (reviewed by Bass, 2012), especially in light of the pathological correlations with perturbed endogenous clock function. Major metabolic organs display rhythmic physiology, which includes cyclic insulin secretion from the pancreas (Perelis et al., 2015), therefore it is perhaps not surprising to find clock disruptions associated with diabetes, obesity, and metabolic syndrome in genetic mouse models (Marcheva et al., 2010, Turek et al., 2005). These associations are alarming given the ubiquity of circadian disruption in modern society, including the high prevalence of shift work and increased light exposure at all hours of the night. Adiposity, weight, lipid and glucose metabolism are the most common alterations found in animal studies modeling disruptions in sleep, light, and feeding (Opperhuizen et al., 2015), although much work remains to delineate impacts of environmental cues on metabolic variations. Specifically, while the impact of light on transcriptional clocks is understood (Emery et al., 1998), metabolic variations due to light exposure have not been directly addressed.

The mechanisms of the core endogenous circadian clock were first described in Drosophila melanogaster, where the core clock consists of a transcription-translation feedback loop very similar in structure to mammalian clocks (Glossop and Hardin, 2002). The CLK/CYC complex binds to E-box elements to drive expression of many gene products, including period and timeless (reviewed in Hardin, 2011). These protein products (PER and TIM) dimerize and translocate to the nucleus to repress CLK/CYC activity. Timely degradation of TIM is followed by turnover of PER, which breaks the repressive arm of the clock and allows CLK-CYC-mediated transcriptions to resume. This clock mechanism exists across most fly tissues, including metabolically active organs such as the fat body (functionally similar to the mammalian liver and

adipose). While flies have served well to uncover the architecture of circadian processes, relatively few studies have utilized flies as a model organism in chronometabolic studies. The interaction of the core clock with metabolic homeostasis has been established through analysis, including comparative transcriptomics, of the fat body (Xu et al., 2011). Oscillations of transcripts in the fat body are controlled by clocks in the fat body or the brain (Xu et al., 2011; Erion et al., 2016; Barber et al., 2016) and most are also influenced by the time of feeding, however the extent of circadian control on metabolism has not yet been established.

Circadian transcriptomics in mice have shown a striking enrichment of oscillating transcripts for rate-limiting metabolic enzymes (Panda et al., 2002). Many of these metabolic enzymes also display ultradian patterns of 8 or 12-hour periods, which may reflect the impact of feeding and behavior on metabolism superimposed on the canonical 24-hour pattern. Circadian and ultradian metabolite cycles have recently been discovered in mouse liver using LC-MS metabolomics with high sampling resolution (Krishnaiah et al., 2017). Metabolomics is an increasingly popular approach to uncover metabolic rhythms through holistic analysis of metabolites by MS or NMR (Rhoades et al., 2017). While informative, previous studies have not incorporated key experimental design considerations in understanding circadian processes. Most circadian metabolomics studies in animals have employed 4 or 6-hour sampling resolution (Abbondante et al., 2015, Chaix et al., 2014, Eckel-Mahan et al., 2013), which is adequate to discern 24-hour periods, but may not be sufficient to discover ultradian metabolic patterns (Krishnaiah et al., 2017). Additionally, in these studies, tissues or blood are harvested under light:dark conditions, which does not allow for separation of light and clock-driven metabolic patterns. True circadian patterns can be ascertained only in constant environmental conditions, such as DD.

In this study, we sought to profile small polar metabolites with 2-hour resolution in Drosophila under both LD and DD. Additionally, *per* mutant flies were analyzed to assess the importance of the genetic clock in driving metabolite rhythms. To provide a broad spectrum of metabolites, we used LC-MS, and chose to analyze fly bodies to reduce confounding effects from

the brain and eyes. Univariate periodicity algorithms and multivariate statistical models were employed to discern both metabolite oscillations under a functional clock and broader metabolic shifts in the context of both light cues and the per mutation. The metabolite rhythms we report here are conserved across species, suggesting that flies will serve an important role in deriving mechanistic understanding of the chronometabolic connection.

3.3 Materials and Methods

Drosophila strains

Both fly strains (*Drosophila melanogaster*), which included Iso^{31} (isogenic w^{1118} stock) and per^{01} mutants (Konopka and Benzer, 1971), were maintained at 25°C in 12:12hr LD conditions on standard cornmeal/molasses medium.

Fly entrainment and collection

Male flies were sorted shortly after eclosion and entrained in LD incubators for a minimum of 3 days before circadian collection, by which all flies were 5-10 days of age. Wild-type flies were either maintained in LD conditions (WT-LD), or placed in constant darkness for a minimum of 24 hours before collection to assess light-independent rhythms (WT-DD), while per^{01} flies were maintained in LD cycles (Per-LD). 12 flies were collected at each time point, in duplicate. Zeitgeber time (ZT) 0 corresponds to lights on, with lights-off at ZT12.

Metabolite extraction and LC-MS/MS metabolomics

Fly heads and bodies were separated before metabolite extraction. Adapted from the Bligh-dyer extraction (Bligh and Dyer, 1959), 600µL of cold 2:1 methanol:chloroform was added to the fly body samples and homogenized in a bead-based tissue homogenizer at 25Hz for 4 minutes (TissueLyser II, Qiagen, Hilden, Germany). 200µL of both water and chloroform was then added, followed by centrifugation at 18787xg for 7 minutes at 4°C. 350µL of the upper layer, comprising the aqueous layer, was separated and dried down overnight under vacuum. Samples

were resuspended in 100µL of 50:50 water:acetonitrile for LC-MS injection. Liquid chromatography conditions and mass spectrometer parameters for HILIC analysis for small polar metabolites were executed as previously reported (Rhoades and Weljie, 2016). 5µL injections were performed for each sample in duplicate on a Waters ACQUITY H-Class UPLC coupled to a Waters TQ-S micro mass spectrometer (Milford, MA), utilizing an XBridge BEH Amide column for chromatographic separation (2.1 x 150mm, 2.5µm). The LC solvents consisted of 95:5 water:acetonitrile with 20mM ammonium acetate at pH 9 (mobile phase A) and acetonitrile for mobile phase B. The gradient was changed from 15 to 70% A over 5 minutes at 0.15 mL/min, followed by an isocratic hold for 10 minutes. The column was washed in 98% A, then re-equilibrated in starting conditions for 5 minutes before the next injection. The MS operated in ion-switching mode with a capillary voltage of 3kV for ESI+ mode and 2kV for ESI-. The desolvation gas flow was set to 900L/Hr and desolvation temperature at 450°C , with the source temperature set at 150°C. Metabolites were detected using MRMs, with mass transitions and voltages optimized as previously described.

Data processing and analysis

LC-MS chromatograms were processed using TargetLynx under MaxxLynx version 4.1. Ion counts were exported and processed in R (version 3.3). QC samples, which consisted of a pooled sample of all biological samples, were injected at the beginning of the batch for LC column equilibration and every 8 injections during the analysis to account for instrumental drift. For every metabolic feature, LOESS was fit to the QC data, which was then used as a normalization factor for the samples as a function of run order. Additionally, metabolic features which appeared in less than 50% of the QC samples and displayed a relative standard deviation (RSD) greater than 30% were dropped from the final dataset. PCA was performed in SIMCA 14.0 (UMetrics, Umeå, Sweden) to observe data integrity through balanced scores and loadings plots. This criteria guided additional data normalization measures, which we addressed through total sample intensity normalization after LOESS correction. Technical replicate values were than averaged

and analyzed for rhythmicity using the ARSER algorithm (Yang et al., 2010) in the MetaCycle R package (version 1.1, Gang et al., 2016). Missing values were imputed through mean-imputation within a single metabolite's observed values, and each circadian time replicate was spread out across 48 hours for rhythmicity testing, which was performed twice through shuffling the replicates to guard against replicate bias. These two ARSER results were then averaged before filtering lists of significant cyclers, testing period lengths of 20-28, 12, and 8 hours.

For block-testing, values in each of the 3 biological groups were further divided into four windows, consisting of values from ZT0-6, ZT6-12, ZT12-18, and ZT18-24. These groups were then tested in a univariate manner, metabolite by metabolite, both across time blocks and biological groups. The resulting p-values from two-sided t-tests were adjusted through a Benjamini-Hochberg False Discovery Rate correction using the stats R package (version 3.3, Benjamini and Hochberg, 1995). For multivariate assessment of global metabolic shifts within and across biological groups, orthogonal partial least squares discriminant analysis (OPLS-DA) models were fit in SIMCA 14.0, with model validity assessed through a cross-validation Q^2 value and CV-ANOVA p-value (Eriksson et al., 2006). Plots were made using VennDiagram v1.6, heatmap3 v1.1, and ggplot2 v2.2.

3.4 Results

Circadian and light-induced rhythms identified in wild-type flies

Rhythmicity tests in WT flies under LD conditions revealed 34 metabolites with period lengths between 20-28 hours, which represents 14.4% of detected metabolites and contains an abundance of sugars and acylcarnitines ($p<0.05$ and $q<0.3$, Figure 3.1A, Table 3.1). Seven metabolites exhibited a 12-hour period and eleven exhibited an 8-hour period in this WT-LD. 29 metabolites yielded true circadian patterns in the WT-DD condition, with seven metabolites overlapping the WT-LD group (fructose, 3-hydroxybutyrate, acetyl-amino sugars, gluconate, an unknown monosaccharide, ribitol/xylitol, riboflavin) (Figure 3.1B, Figure 3.1C). We were thus able to identify sets of metabolite rhythms driven by endogenous clocks, LD cycles, or both.

Ultradian metabolites enriched in *per* mutants

No circadian metabolites, nor ultradian metabolites of 8-hour periods, were detected in *per* flies, however we noted an abundance of metabolites with 12-hour periods (Table 3.1). Interestingly, the amplitudes for these cyclers were significantly higher than 12-hour cyclers in the WT-LD condition (p=0.001, Figure 3.2). While no overlapping metabolites arose across 12-hour cyclers in WT-LD and Per-LD conditions, ten out of fourteen 12-hour cyclers in the Per-LD condition were found in the 20-28 hour list for WT-LD (2-oxovalerate, fructose, acetyl-amino sugars, gluconate, ribitol/xylitol, inositols, N6-acetyllysine, cAMP, acetylcholine, and aminobenzoates/trigonelline), and cGMP additionally overlapped with WT-DD (Figure 3.3). Thus, of these fourteen 12-hour cyclers, only three did not also oscillate in WT-LD or WT-DD (imidazole, bisphosphoglycerate, aconitate). Four metabolites overlapped across all three lists; these consisted only of carbohydrates (fructose, acetyl-amino sugars, ribitol/xylitol, gluconate) and may reflect a consequence of complex interactions of light and clock-driven feeding behavior.

Unique phases in significant metabolites across experimental groups

Of the significant cyclers with a 20-28 period in WT-LD, we noted major phases of metabolites in the morning (ZT4-6) and at the end of the active phase (ZT12). Peak expression of cycling metabolites in DD was, on the other hand, largely condensed towards the end of the active phase (Figure 3.4A). The abundance of 12-hr cyclers in Per-LD also had highly condensed phases, with antiphasic peaks in the early light and dark phases and an absence of the evening rush seen in the WT-LD (Figure 3.4B). Additionally, the seven metabolites that overlap with WT-LD and WT-DD display dramatically different phases under these two conditions and is most strikingly the case for the sugar compounds (Figure 3.4C).

Light and clock driven alterations in global metabolite profiles

As expected, the *per* mutants displayed large global metabolite differences from WT, irrespective of time, through principal components analysis, while WT-LD and WT-DD samples had general overlap in their metabolic signatures (Figure 3.5). To gauge global time-dependent shifts in metabolite signatures, the time points were binned into four blocks of ZT0-6, ZT6-12, ZT12-18, and ZT18-24. Using univariate testing, the metabolic separation between WT-LD and WT-DD could only be found during the ZT6-12 window ($p<0.05$ and $q<0.3$, Table 3.2), suggesting a set of metabolites that are driven by light. Within the WT-LD group, strong diurnal patterns were noted in amino acid and nitrogen metabolism (Figure 3.6A), while medium- to long-chain acylcarnitines comprised the majority of time-dependent metabolite changes in WT-DD (Table 3.2). Differences between the WT-LD and Per-LD groups were noted across each time block, driven mostly by changes in amino acid and nitrogen metabolism (Figure 3.6B). These diurnal pathway enrichments in WT-LD were lost in the *per* flies, as no metabolites were found to significantly change within the *per* dataset across the blocked times. As an additional measure of global metabolite shifts, OPLS-DA scores plots demonstrated a convergence of WT-LD and WT-DD profiles during the dark phase (Figure 3.7). The strongest discriminant model with the most predictive power across the three groups occurred at ZT6-12 ($R^2X=0.585$, $R^2Y=0.874$, $Q^2=0.654$, $p=0.003$), while a significant model could not be fit at ZT18-24 ($p=0.24$).

Conserved rhythms across published circadian metabolomic datasets

We compared our list of significantly diurnal metabolites from WT-LD to the only currently published fly metabolomics study, which analyzed small polar metabolites in whole flies under conditions of cycling light and temperature, and found overlaps in alanine, tryptophan, AMP, creatine, and lactate (Gogna et al., 2015). Of these metabolites, the phases for alanine and creatine aligned closely across studies (ZT15 and ZT18 for creatine, ZT13 and ZT16 for alanine), however the phases of other metabolites generally differed.

For a comparison of significantly circadian metabolites conserved across species, we separated datasets that utilized LD and constant darkness paradigms (Table 3.3). A majority of

prior circadian metabolomics studies have employed LD, and we found 15/35 of the WT-LD 20-28 hour cyclers in our study to oscillate in other species, including zebrafish, mice, and humans (Li et al., 2015, Zwigahft et al., 2015, Giskeodegard et al., 2015, Davies et al., 2014, Chaix et al., 2014, Hatori et al., 2012, Eckel-Mahan et al., 2012, Dallmann et al., 2012, Ang et al., 2012). These metabolites include glutamate (found in all 4 species), 2-hydroxyglutarate, 2-aminoadipate, fructose, 3-hydroxybutyrate, 4-guanidinebutanoate, threonine, citrate/isocitrate, carnitine, inositol, hydoxyhexadecenoylcarnitine, aminobenzoates/trigonelline, acetylcarnitine, UDP-glucose, and riboflavin. Only one other study has performed high-resolution small polar metabolite analysis under constant darkness in mice (Krishnaiah et al., 2017), which yielded 3-hydroxybutyrate, adenine, and fructose as overlaps to this study. However, many of these studies in mice and humans sample liver and blood respectively, which will likely differ from systemic metabolic rhythmicity in the whole fly body.

3.5. Discussion

To our knowledge this is the first report of high resolution circadian metabolomics which tested key hypotheses regarding the separation of circadian and environmentally-driven metabolic processes using both genetic clock mutants and constant darkness paradigms. First, we discovered a unique set of metabolites that continue to oscillate in constant darkness, clarifying the complex interactions of our endogenous clocks and environmental cues. In our experimental design, DD sampling did not occur until the flies were exposed to a minimum of 24 hours of darkness, which may have dampened the rhythms but allowed further separation of metabolite patterns from the impact of light. We suspect that, given many of the overlapping metabolites with 20-28hr periods across LD and DD were carbohydrates, a continued feeding rhythm is a major driver in metabolite rhythms, as has been noted in proteomics profiles in mouse liver (Mauvoisin et al., 2014). These sugars were uniformly increased during the light phase in WT-LD compared to WT-DD, perhaps as a result of a heightened feeding rhythm under the presence of light. In addition, we demonstrate a set of light-induced metabolites with large

enrichments in amino acid and nitrogen metabolism, which is likely correlated to light-induced locomotor behavior and overlaps with a large enrichment of amino acid rhythms in mice (Minami et al., 2009). Interestingly, indoleacetic acid and kynurenine were among the few metabolites that decreased in the presence of light, suggesting an effect of light on dampening certain neurotransmitter processes. Only one study previously discovered 14 diurnal polar metabolites with 2-hour resolution in whole adult flies using NMR (Gogna et al., 2015), with one additional experiment employing lipid profiling at 4-hour resolution (Katewa et al., 2016). Gogna et al. sought to maximize metabolite rhythmicity by entraining flies in both LD and temperature cycles, which may establish a control baseline metabolite rhythm, but is insufficient to discern endogenous circadian processes. Despite adequate sampling resolution, this study did not employ statistical algorithms to discern periodicity, such as JTK_CYCLE or ARSER (Hughes et al., 2010, Yang and Su, 2010). We note some overlaps of rhythmic but non-circadian metabolites, however the phases generally differed, and the relative insensitivity of NMR compared to LC-MS yielded fewer possible comparisons. Given the impact of environmental cues on metabolite phases and rhythmicity noted here, the relatively low overlap with Gogna et al. is perhaps not surprising, and future studies will need to tease out the relative contributions of additional environmental cues such as temperature and humidity. Additionally, this report analyzed whole flies, rather than bodies in our study. Contributions of metabolites in the eye and brain may impart some additional effects to the detected metabolome.

We were surprised to find how many metabolites uniquely oscillated in WT-DD compared to WT-LD. We can partly attribute this disparity to the fact that six metabolites did not meet our significance threshold ($0.05 < p < 0.10$) in the WT-DD but were significant in WT-LD; additional work will be required to separate statistical power versus biological variance. Within the WT-DD group, diurnal patterns in acylcarnitines were observed, which corroborates nicely with some existing metabolomics studies in humans under constant or highly controlled conditions (Ang et al., 2012, Dallmann et al., 2012, Davies et al., 2014). We speculate that under constant environmental conditions, rhythms in lipid metabolism may persist but are largely transformed

under the presence of light, and may partly explain our set of unique rhythms in constant darkness. Previous transcript analyses in Drosophila heads have also noted close to half of significantly cyclic transcripts in DD were not found in LD conditions (Wijnen et al., 2006). This alignment with our metabolomic findings suggests transcriptional and metabolic programs are largely transformed in the presence of light cycles. Interestingly, this study also uncovered a similarly large number of cycling transcripts in clock mutants under LD relative to WT flies in LD, further demonstrating the substantial impact of light on rhythmic processes. However, the light-driven transcripts displayed 24-hour periods instead of the 12-hour enrichment we report here for metabolites, which may be a consequence of statistical testing in 2-hour versus 4-hour sampling resolution. We can confidently speak to the impact of light on the global metabolome given that metabolite profiles were indistinguishable during the dark phases of the WT-LD and WT-DD groups by both univariate and multivariate testing.

One drawback in LC-MS is relatively high variance, which may explain a rhythmicity of over 14% of detected metabolites in our study, compared to upwards of 50%, as reported in mouse liver (Krishnaiah et al., 2017). However, given peripheral clocks tend to oscillate with unique phases (Zhang et al., 2014), we would not expect such a high enrichment of circadian metabolites in whole bodies. In addition, the age of flies in our study ranged from 5-10 days, which may add some additional variance in the dataset. The mouse liver metabolomics study incorporated 1-hour sampling resolution, which greatly augments power in circadian testing algorithms, even over the 2-hour resolution employed here, and may additionally explain relatively lower enrichment in cyclic metabolites. Given our experimental design and biological variance, we felt ARSER analysis was more suitable to discern periodicity in the data, although it has been reported to yield a higher false-positive rate for transcriptomics analyses (Wu et al., 2016). Future experiments may require higher sampling density and/or replicates, while remaining mindful of resource allocation for large-scale LC-MS experiments (Dunn et al., 2011).

The abundance of 12-hour cyclers in Per-LD likely derive from light and activity patterns, which would be otherwise dampened or overridden by an intact clock and may in part explain the

strong amplitudes of these ultradian rhythms. A dampened secondary peak of feeding in *per* mutants (Barber et al., 2016) may explain the uniquely biphasic phase enrichments of circadian metabolites in WT-LD compared to a strong enrichment of antiphasic phases for Per-LD at ZT4 and ZT16 in Figure 3.4. Previous transcriptomics studies found large phase enrichments in WT flies at ZT8 and ZT20, which differ slightly from metabolite phases (Ceriani et al., 2002). For WT flies in constant darkness, the feeding rhythm loses much of its biphasic nature (Barber et al., 2016), which would corroborate our speculation that the second feeding peak persists without light and accounts for the condensed metabolite phases towards the end of the subjective day in WT-DD. These changes in feeding patterns are consistent with alterations in locomotor activity, which also tends to lose bimodality in DD (Helfrich-Förster, 2000). The large overlap of Per-LD 12-hour cyclers and circadian WT-LD metabolites suggest a complex synergy of circadian and light-driven processes, with a conserved surge of metabolite phases in the morning. We posit that while clocks serve an evolutionary advantage in anticipating environmental cues like light, these cues can themselves trigger adaptive responses in activity and feeding, which would confer an additional survival advantage in for e.g. seasonal changes of LD cycles.

Drosophila can serve as a practical model to develop mechanistic detail from these metabolomics studies, with the depth of genetic tools, ease of environmental manipulations, low cost, and high throughput experimentation. Encouragingly, some conserved circadian metabolites were found across multiple species, potentially expanding the utility of Drosophila as a powerful model for chronometabolic studies. A sizable number of human circadian and sleep metabolomics studies have been performed, which has yielded some conserved metabolite hits even across analytical platforms (Rhoades et al., 2017). We hope expanded experimental designs for flies will yield a similar result.

Acknowledgements

We thank Paula Haynes for assistance in Drosophila husbandry and Gang Wu and Michael Hughes for insight into algorithms for circadian rhythmicity testing. S.D.R. is supported

through a Pharmacology T32 Training Grant (T32 GM008076), and additional support for this work comes from the National Center for Research Resources (Grant No. UL 1RR024134).

Table 3.1 Significant metabolites (p<0.05, q<0.3) from ARSER results for each group and period

WT-LD 20-28hr	WT-LD 12hr	WT-DD 20-28hr
Glutamate	Adenine	cGMP
2-Hydroxyglutarate	2-Aminopimelic Acid	Hydroxytetradecanoylcarnitine
Imidazoleacetic Acid	3-Indoleacetic Acid	Hydroxybutrylcarnitine
2-Aminoadipate	Unknown_360.50->85.03_7.04	Lipoamide
2-oxovalerate	Disaccharide	LPC 20:1
Fructose	Thiamine	Riboflavin
3-Hydroxybutyrate	Hydroxytetradecanoylcarnitine	Fructose
Hydroxydecenoylcarnitine	**WT-LD 8hr**	Unknown_154.07->137.01_4.11
Acetyl-amino Sugars		Unknown_516.30->184.00_6.07
4-Guanidinobutanoic Acid	Cystathionine	Gluconate
Gluconate	Unknown_342.50->85.03_7.00	alpha-Ketoglutarate
Threonine	Unknown_137.03->93.95_4.10	Phenylalanine
Unknown_358.50->85.03_4.58	Unknown_203.15->70.07_6.78	Lauroylcarnitine
Cytosine	Xanthine	Acetyl-amino Sugars
Citrate/Isocitrate	Unknown_152.00->110.00_5.76	Unknown_152.00->135.00_5.27
Unknown_166.05->74.02_5.93	CDP-Choline	**WT-DD 12hr**
Unknown_198.00->181.00_5.29	Oxypurinol	
Carnitine	Unknown_112.00->95.00_7.60	Spermine
Ribitol/Xylitol	Citrulline	**WT-DD 8hr**
Inositols	Lysine	
Asparagine	**WT-DD 20-28hr**	Hexose-6-Phosphates
N6-Acetyllysine		**Per-LD 12hr**
Hydroxyhexadecenoylcarnitine	Stearoylcarnitine	
cAMP	Ribitol/Xylitol	cAMP
Dodecenoylcarnitine	3-Indoleacetic Acid	Fructose
Acetylcholine	Acetyl-CoA	N6-Acetyllysine
Hydroxytetradecenoylcarnitine	Tryptophan	Gluconate
SAM	Unknown_137.03->93.95_4.10	Aminobenzoates/Trigonelline
Riboflavin	3-Hydroxybutyrate	cGMP
Aminobenzoates/Trigonelline	Choline	Acetylcholine
Itaconic Acid	NADPH	Imidazole
UDP-Glucose	Deoxyguanosine	Acetyl-amino Sugars
N-Acetyl-L-Tyrosine	Deoxyinosine	2-oxovalerate
Acetylcarnitine	Adenine	Inositols
	NADH	Bisphosphoglycerates
	Unknown_198.00->181.00_5.29	Aconitate
		Ribitol/Xylitol

Table 3.2 Univariate metabolite comparisons between WT-LD and WT-DD or within WT-DD across time. Direction refers to either relative to WT-LD, or relative to the first of two time blocks within WT-DD

WT-LD to WT-DD Comparison

Metabolite	Comparison	P-Value	Q-Value	Direction
Dimethylarginine	ZT6-12	3.78E-05	0.008929	↓
Unknown_282.12->150.00_6.38	ZT6-12	0.000187	0.022123	↓
Threonine	ZT6-12	0.000789	0.062044	↑
Deoxyguanosine	ZT6-12	0.005291	0.228265	↑
Unknown Monosaccharide	ZT6-12	0.006958	0.228265	↑
3-Indoleacetic Acid	ZT6-12	0.006966	0.228265	↓
Asparagine	ZT6-12	0.007666	0.228265	↑
Fructose	ZT6-12	0.009898	0.228265	↑
Ribitol/Xylitol	ZT6-12	0.012777	0.228265	↑
Cytosine	ZT6-12	0.014432	0.228265	↑
Unknown_184.07->86.10_4.22	ZT6-12	0.015219	0.228265	↑
Aminobenzoates/Trigonelline	ZT6-12	0.015475	0.228265	↑
N6-Acetyllysine	ZT6-12	0.01598	0.228265	↑
Gluconate	ZT6-12	0.016001	0.228265	↑
2-Aminoadipate	ZT6-12	0.016172	0.228265	↑
Kynurenine	ZT6-12	0.016193	0.228265	↓
2-oxovalerate	ZT6-12	0.016443	0.228265	↑
cGMP	ZT6-12	0.022034	0.274692	↑
3-Hydroxybutyrate	ZT6-12	0.022115	0.274692	↑

Within WT-DD Comparison by Time

Acetyl-CoA	ZT0-6 to ZT6-12	0.000114	0.026987	↑
Unknown_282.12->150.00_6.38	ZT0-6 to ZT6-12	0.000474	0.055991	↑
Hydroxyhexadecenoylcarnitine	ZT12-18 to ZT18-24	0.001167	0.151773	↑

Unknown_342.50->85.03_7.00	ZT12-18 to ZT18-24	0.001494	0.151773	↑
Stearoylcarnitine	ZT12-18 to ZT18-24	0.001929	0.151773	↑
Lauroylcarnitine	ZT12-18 to ZT18-24	0.003638	0.175813	↑
Hydroxytetradecanoylcarnitine	ZT12-18 to ZT18-24	0.004489	0.175813	↑
Hydroxyhexadecanoylcarnitine	ZT12-18 to ZT18-24	0.00517	0.175813	↑
cis-5-Tetradecenoylcarnitine	ZT12-18 to ZT18-24	0.005835	0.175813	↑
2-Aminoadipate	ZT12-18 to ZT18-24	0.00596	0.175813	↓
Unknown_245.08->113.03_5.56	ZT12-18 to ZT18-24	0.007294	0.191273	↓
Glutamate	ZT12-18 to ZT18-24	0.009185	0.216764	↓
Xanthine	ZT12-18 to ZT18-24	0.012658	0.271562	↑

Table 3.3 Circadian metabolite hits across studies under WT-LD paradigm

Metabolite	**Overlapping Studies**
Glutamate	Fly1, Zebrafish1, Mouse4, Human3
2-Hydroxyglutarate	Fly1, Mouse4
2-Aminoadipate	Fly1, Mouse2, Mouse4
Fructose	Fly1, Mouse2, Human3
3-Hydroxybutyrate	Fly1, Mouse4
4-Guanidinobutanoic Acid	Fly1, Mouse5
Threonine	Fly1, Zebrafish1, Mouse2
Citrate/Isocitrate	Fly1, Mouse4, Human3
Carnitine	Fly1, Zebrafish1
Inositol	Fly1, Mouse4, Mouse5
Aminobenzoates/Trigonelline	Fly1, Mouse4
Acetylcarnitine	Fly1, Mouse2, Human4
UDP-Glucose	Fly1, Mouse2
Riboflavin	Fly1, Mouse2
Adenine	Fly1, Mouse4
Citrulline	Fly1, Mouse2, Mouse4
Lysine	Fly1, Fly2, Mouse2, Mouse4, Human4
Xanthine	Fly1, Mouse2

Legend

Fly1 - Rhoades
Fly2 - Gogna
Zebrafish1 - Li
Mouse1 - Zwighaft
Mouse2 - Abbondante
Mouse3 - Chaix
Mouse4 - Hatori
Mouse5 - Eckel-Mahan
Human1 - Giskeodegard
Human2 - Davies
Human3 - Dallmann
Human4 - Ang

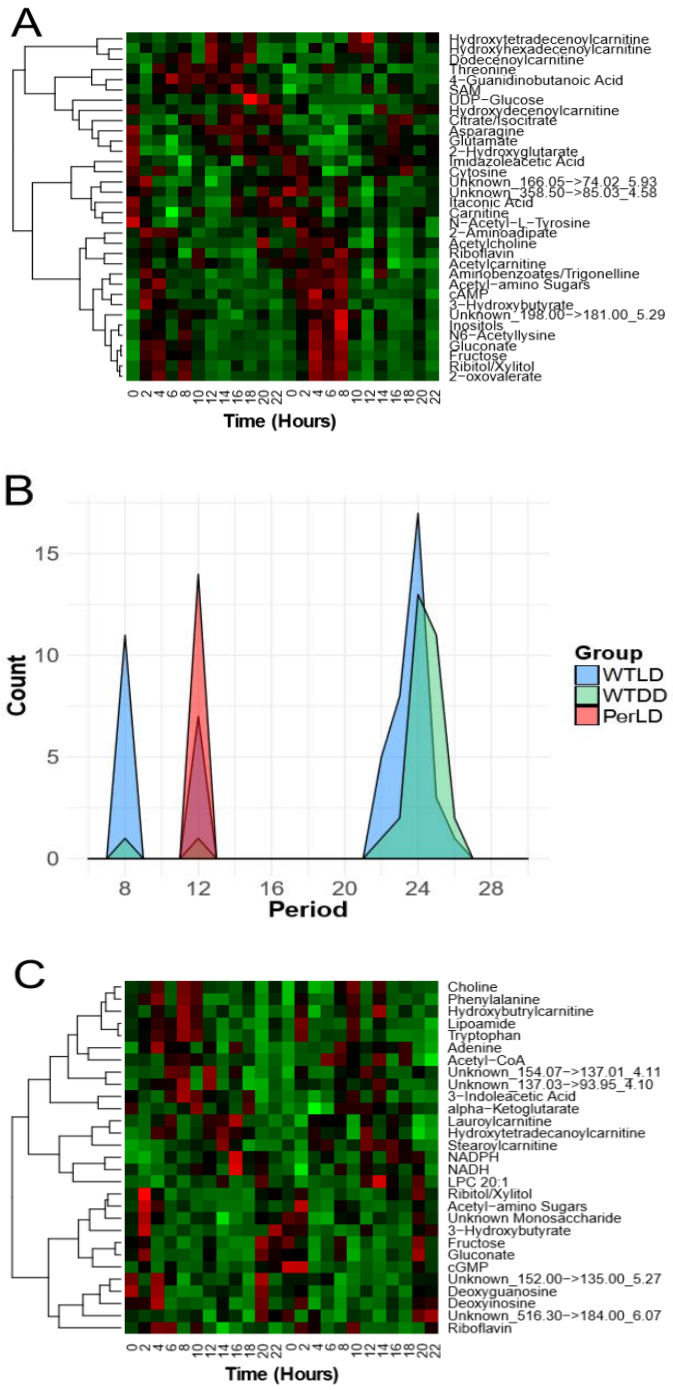


Figure 3.1 Identification of metabolites expressed rhythmically in fly bodies. (A) Metabolites that significantly cycle with 20-28hr periods in WT-LD ($p<0.05$, $q<0.3$, ARSER algorithm). (B) Metabolites that significantly cycle with 20-28hr periods in WT-DD ($p<0.05$, $q<0.3$, ARSER algorithm). (C) Density of circadian periods by group, whereby ARSER testing occurred at 8, 12, or 20-28hr searches.

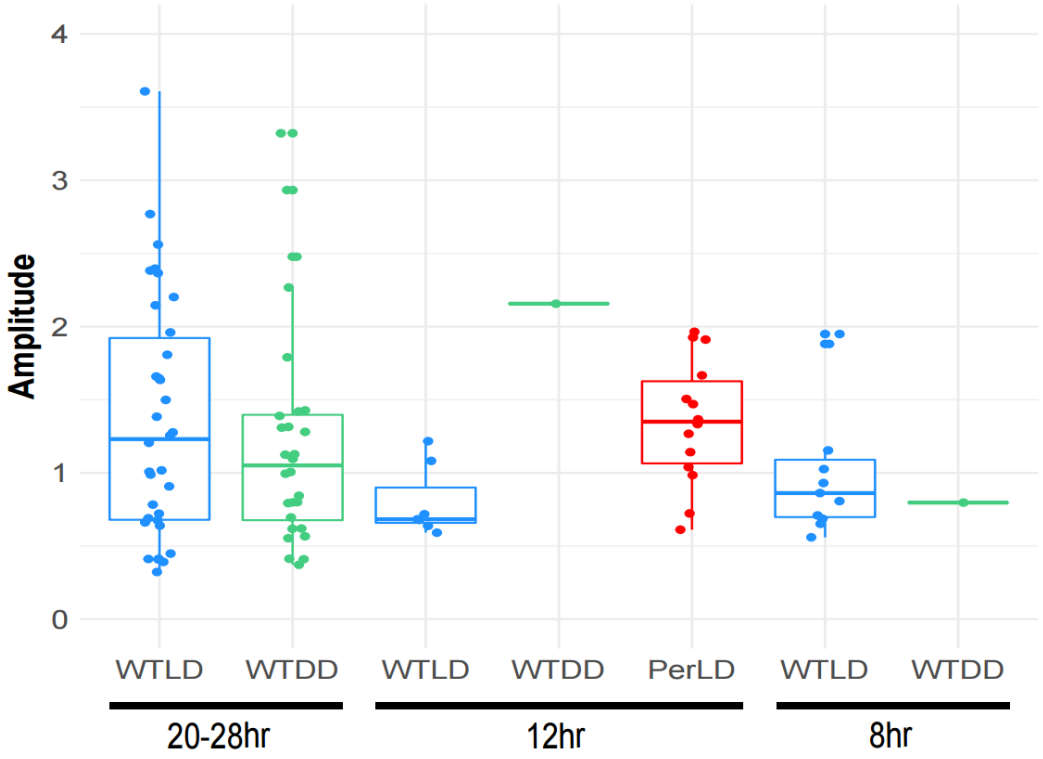


Figure 3.2 Amplitudes for each significant metabolite within each tested period and group. Amplitude values were derived from ARSER calculations. Per-LD 12hr cyclers displayed a significantly higher fold change than WT-LD 12hr cyclers (p=0.001).

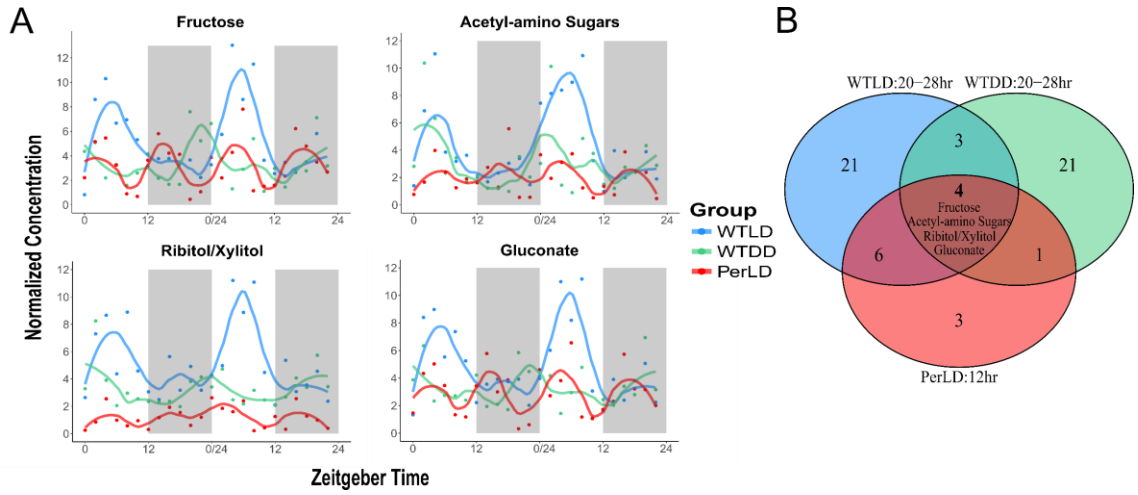
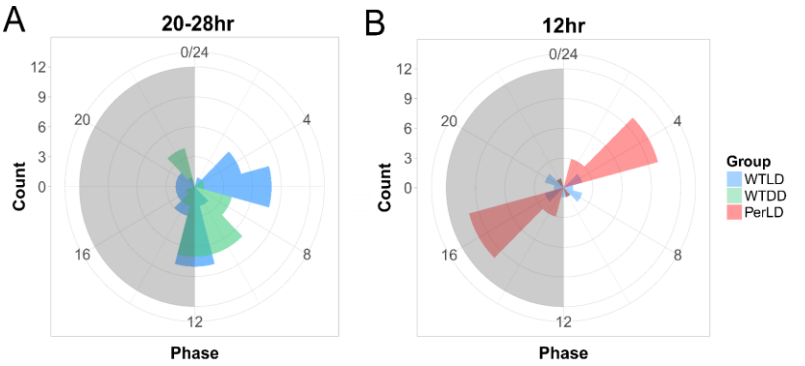


Figure 3.3 Effects of light and the clock on the pattern of cycling metabolites (A) Time-course concentrations of the four carbohydrate species which oscillate with 20-28hr periods in WT and a 12hr period in *per*. (B) Overlapping cycling metabolites across all three groups with unique periods.



C

	WT-LD Phase	WT-DD Phase
3-Hydroxybutyrate	4	12
Acetyl-amino Sugars	4	14
Fructose	6	23
Gluconate	6	23
Ribitol/Xylitol	6	13
Riboflavin	6	5
Unknown Monosaccharide	7	13

Figure 3.4 Radial plot indicating phases for significant 20-28hr cyclers in WT-LD and WT-DD (A) and 12hr cyclers in all three groups (B), as calculated by ARSER. C. Phases of the seven circadian overlaps in WT-LD and WT-DD.

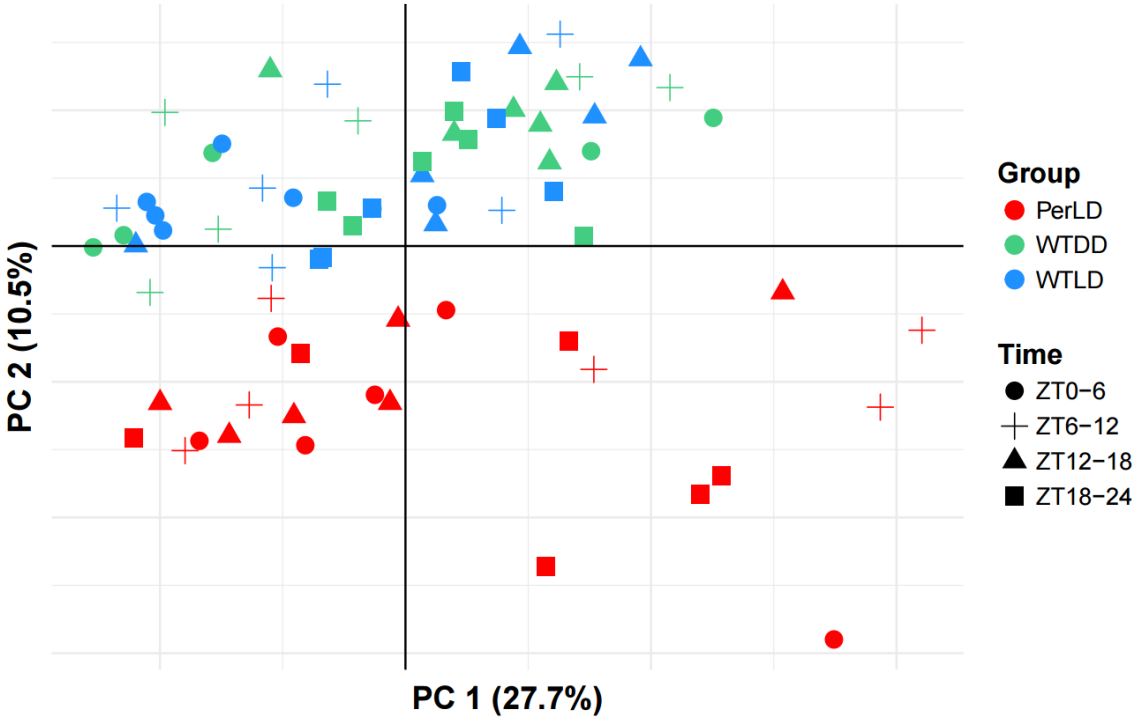


Figure 3.5 Principal components scores plot for all samples. Variance explained for the first two components were 27.7% and 10.5% respectively.

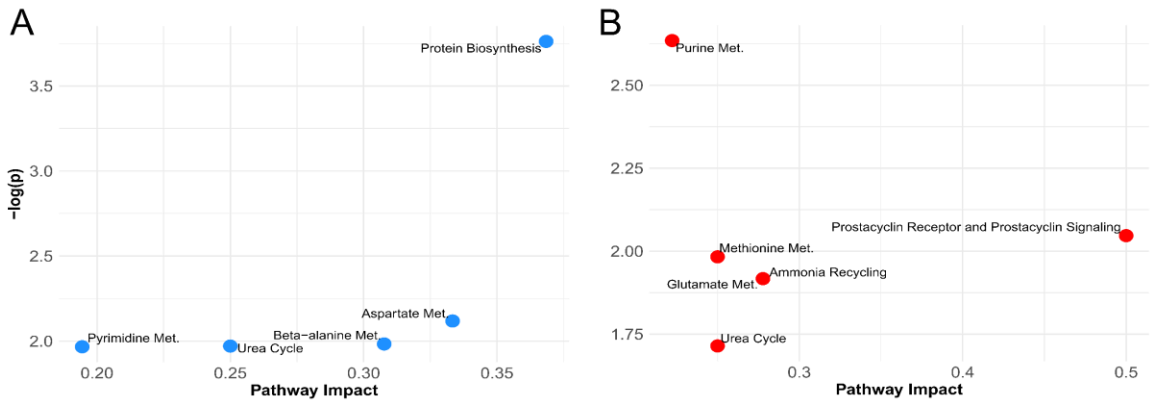


Figure 3.6 (A) Metabolite set enrichment analysis for metabolites which altered significantly in WT-LD across time-block testing ($p<0.05$, $q<0.3$) (B) Metabolite set enrichment analysis for metabolites which altered significantly across WT-LD and Per-LD, considering all four time-blocks.

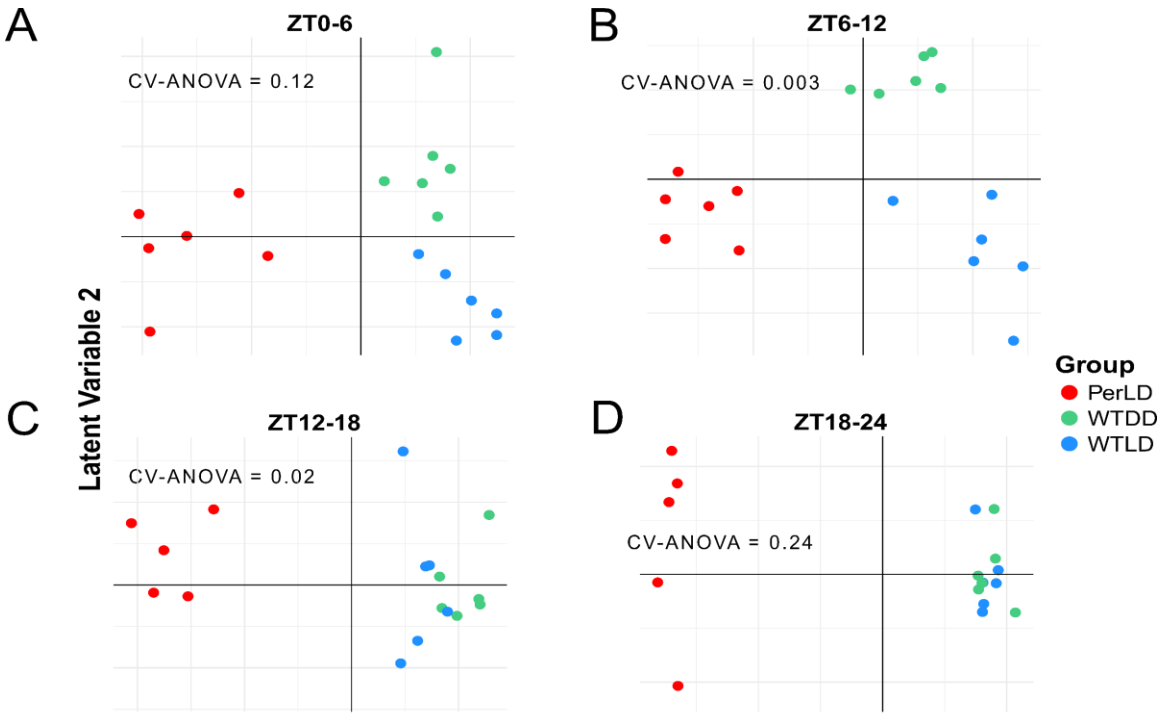


Figure 3.7 OPLS-DA scores for discriminant analysis of all three groups in each of the four time windows. Significant models were fit to ZT6-12 (B, R^2X=0.585, R^2Y=0.874, Q^2=0.654, p=0.003) and ZT12-18 samples (C, R^2X=0.458, R^2Y=0.488, Q^2=0.265, p=0.02) but not ZT0-6 (A, p=0.12) nor ZT18-24 (D, p=0.24), noting the convergence of metabolite profiles in WT-LD and WT-DD at the end of the dark phase in ZT18-24.

CHAPTER 4 – *In vivo* circadian isotopolomics platform uncovers unique diurnal flux patterns perturbed under sleep deprivation

4.1 Abstract

Basic research in circadian rhythms has recently benefitted from a diversity of high-throughput systems-level approaches, however they invariably describe snapshots of a dynamic process. Furthermore, the metabolic connections to sleep are largely undefined, and new approaches are necessary to unearth the impact of sleep perturbation on diurnal reaction kinetics. Here, a novel *in vivo* platform is developed to detect stable isotope labeling patterns in *Drosophila* across circadian time. The impact of perturbed sleep/wake cycles on metabolic flux is observed using *fumin* flies, a genetic model of hyperactivity and sleep deprivation. Microcapillary injections of $^{13}C_6$ glucose into flies at a safe and tolerable dose provide a unique method to study acute metabolic flux in downstream oxidative pathways using bespoke liquid-chromatography mass spectrometry (LC-MS) metabolomics methodology. Carbon tracing from glucose into serine displayed a robust circadian rhythm, which is lost under sleep deprivation. Glutamine labeling patterns, likely reflective of TCA cycle activity, demonstrated an ultradian cycle in wild-type flies but gained circadian rhythmicity in *fumin* mutants. Labeled glutathione displayed a striking 8-hour ultradian pattern that is lost under sleep deprivation, which may stem from altered redox defense mechanisms in response to perturbed sleep/wake cycles. Global changes in isotopolome patterns in wild-type flies were most noticeable in the morning hours, suggesting a catabolic rush to meet the requirements of energy production during the active phase. Herein a new means of analyzing circadian flux is presented, overcoming limitations in interpretability of chronometabolic processes from prior 'omic analyses.

4.2 Introduction

Research in circadian rhythms and sleep has recently benefitted from a diversity of systems-level approaches, including transcriptomics, proteomics, and metabolomics (Zhang et al., 2014, Mauvoisin et al., 2014, Krishnaiah et al., 2017, Rhoades et al., 2017). Metabolite concentrations are particularly promising to derive the phenotypic outputs of the clock and sleep status, as metabolites are viewed as products of active physiological processes, and interpretations from other 'omics technologies can be confounded by the variable lag between transcription and translation of a given gene product and circadian cycles in post-translational and epigenetic modifications (Krishnaiah et al., 2017, Lamia et al., 2009, Feng et al., 2011). However, while metabolites may couple more closely to organismal phenotypes than other systems-level approaches, any of these techniques are inherently snapshots of a dynamic system. Much like taking a picture of a highway versus measuring the cars' velocities, static metabolic profiling may give clues into differential pathway activities, but cannot be used to truly ascertain metabolic flux (Zamboni et al., 2015). Flux is a more cogent assessment of the metabolic phenotype, and is highly conserved across species to maintain enzymatic efficiency under thermodynamic and osmotic constraints (Park et al., 2016). Despite the inherently transitory nature of circadian and sleep processes, few studies have focused on analyzing the rates of metabolic reactions or pathways with respect to time. New approaches must be employed to add this kinetic dimension to our understanding of the chronometabolic connection.

Metabolic flux cannot be directly observed in an active biological system, however one common approach to infer flux is to analyze metabolism of stable isotope tracers to downstream products over specified time intervals. The objectives in isotope-tracer experimentation have largely shifted from biochemical discovery in the mid-20th century towards characterizing metabolic phenotypes, particularly in recent comparative analyses of cancer metabolism. Observation of the Warburg effect has driven numerous isotope labeling studies which incorporate glucose and analyze isotope enrichments in the pentose phosphate pathway (PPP) and TCA cycle (Fan et al., 2014, Fan et al., 2013). Some tracer studies have been recently

extended to one-carbon metabolism and amino acid and lipid synthesis as ancillary metabolic networks become increasingly apropos to disease phenotypes (Ducker et al., 2016, Munger et al., 2008), which are now also understood as important pathways in sleep and circadian physiology (Rhoades et al., 2017). However, no sleep and circadian studies have studies have employed isotope tracing analysis to corroborate static measurements of metabolites and transcripts in these pathways. In addition, labeling studies are typically performed *in vitro*, given the costs and complexity of *in vivo* experiments in mice and humans (Fan et al., 2011). Realistically, sleep and circadian processes are a complex integration of internal oscillators with external environmental cues, including light and feeding, which cannot be adequately modeled *in vitro*. While quantitative flux modeling is considerably more difficult *in vivo*, and currently infeasible in non-steady state eukaryotic systems (Wiechart and Noh, 2013), comparative analysis of qualitative labeling patterns across experimental conditions may still yield fruitful insight into metabolic phenotypes, while maintaining appreciable temporal resolution to discern metabolic cycles.

Drosophila are a boon to sleep and circadian research given their classical sleep characteristics, highly conserved genetic clock mechanism, and relative ease of high-throughput genetic screens (Hendricks et al., 2000). While metabolomic analyses in flies are sparse, many major nutrient sensing pathways and insulin signaling mechanisms found in mammals are also found in flies (Geminard et al., 2009, Fernandez et al., 1995). We therefore reasoned that flies would serve as a good model to assess dynamic metabolic processes across circadian time and under sleep deprivation. Previous studies have added isotope tracers to fly media (Musselman et al., 2013, Katewa et al., 2016), however these studies cannot discern flux at the acute timescale required for circadian analysis. Fortunately, these previous experiments informed us of the tradeoffs in our experimental design. Quantitative flux modeling in eukaryotes typically assumes isotopic equilibrium (Wiechert and Nöh, 2013), which was not obtained in flies until 24 hours of feeding (Musselman et al., 2013). This observation restricted our objectives to gauging relative flux changes acutely across circadian time, as quantitative flux modeling would not be feasible.

Nonetheless, our hypothesis is that we could develop a new platform for isotope tracing in flies, which would uncover relative changes in both core and secondary metabolic activity downstream from glucose. In addition, the metabolic consequence of sleep loss around the circadian cycle is assessed by comparison to a *fumin* mutant fly strain, an established genetic model of sleep loss through dopamine-driven hyperactivity. In these experiments, new tracer incorporation mechanisms are developed and validated, existing LC-MS methods are expanded to analyze numerous labeling possibilities in amino acid metabolism, and data processing pipelines are built to deal with both high-dimensionality datasets and periodic patterns common in large-scale circadian metabolomics experiments (Rhoades et al., 2017).

4.3 Materials and Methods

Drosophila strains

Both fly strains (*Drosophila melanogaster*), which included Iso31 (isogenic w1118 stock) flies and fumin mutants as characterized previously (Kume et al., 2005), were maintained at 25°C in 12:12 LD conditions on standard cornmeal/molasses medium.

Entrainment and tracer injection design

Within one day after eclosion, 15 males were sorted into a new vial for entrainment to LD rhythms for a minimum for 3 days. All flies were between 5-7 days old at the time of injection. One vial was taken every four hours under LD conditions (ZT0, 4, 8, 12, 16, 20) and placed on ice for anesthetization 15 minutes prior to injections. The injection solution consisted of 1M $^{13}C_6$ glucose with a 50x dilution of a blue dye (McCormick, FD&C Blue Dye No. 1), for visualization of uptake and signal normalization, in PBS and injected into the thorax using glass capillaries (3.5” x 1.14mm diameter, Drummond Scientific, Broomall, PA). After injection flies were moved to empty vials without food and placed back in the appropriate light or dark 25°C incubator to metabolize the glucose tracer. After one hour, flies were snap-frozen before metabolite extraction. Five

separate days of injections were performed as biological replicates for both WT and *fumin* mutants. Any flies which died after the injections were discarded.

Geotaxis and locomotion assays

For geotaxis climbing assays, after three days under LD entrainment, 5-7 day-old WT or fumin flies were anesthetized on ice for 15 minutes at ZT0 or ZT12, followed by a control (no injection), PBS, or 1M glucose injection. Flies were then placed in new vials without food for one hour. Climbing activity was measured by timing an individual fly's time to climb 4cm, measured in triplicate. 5-8 flies were used per group. For locomotion assays, the same procedure was taken, except flies were placed in 5 x 65mm glass tubes with 5% sucrose immediately after injections and monitored under Activity Monitoring System devices (DAMS) from Trikinetics (Waltham, MA) for a minimum of 24 hours in 25°C LD incubators. Locomotion data was analyzed using custom MATLAB (Mathworks, Natick, MA) scripts (Yuan et al., 2006), and plotting and statistics were performed in R (version 3.3). Comparisons were made using a Student's one-sided or two-sided t-test when appropriate with FDR adjustments for multiple testing in the locomotion data, using $p<0.05$ and $q<0.2$ as a threshold for significance (Benjamini and Hochberg, 1995). One-sided tests were performed with an assumption that injections would decrease locomotion activity or climbing rate compared to sham control.

Metabolite extraction and LC-MS/MS measurements

Fly heads and bodies were separated before metabolite extraction. Adapted from the Bligh-dyer extraction (Bligh and Dyer, 1959), 600µL of cold 2:1 methanol:chloroform was added to the fly samples and homogenized in a bead-based tissue homogenizer at 25Hz for 4 minutes (TissuLyser II, Qiagen, Hilden, Germany). 200µL of both water and chloroform was then added, followed by centrifugation at 18787xg for 7 minutes at 4°C. 350µL of the upper layer, comprising the aqueous layer, was separated and dried down overnight under vacuum. Samples were resuspended in 80µL of 50:50 water:acetonitrile for injection onto the mass spectrometer. Two

LC-MS/MS analyses were performed to consider both positive and negative ionizing metabolites in ESI. Chromatographic separations for both methods were performed on a Waters ACQUITY H-Class UPLC coupled to a Waters TQ-S Micro MS (Milford, MA), utilizing an ACQUITY UPLC BEH Amide column (2.1 x 150mm, 1.7µm). LC conditions for the positive-ionizing method were performed as described previously (Rhoades and Weljie, 2016), where the solvents consisted of 95:5 water:acetonitrile with 2mM ammonium acetate and 0.2% formic acid (mobile phase A), and 90:10 acetonitrile:water with 2mM ammonium acetate and 0.2% formic acid (mobile phase B). The gradient was changed from 100 to 20.6 % B over 15 min at 0.35 mL/min, followed by a wash of 100 % A for 5 min. Mobile phase B was changed from 0 to 100 % from 20 to 22 min and held for column equilibration until 30 min. For the negative-ionizing method, solvents consisted of 95:5 water:acetonitrile with 20mM ammonium bicarbonate, pH 9 (mobile phase A) and 90:10 water:acetonitrile with 20mM ammonium bicarbonate, pH 9 (mobile phase A). The gradient was changed from 100 to 20.6 % B over 15 min at 0.4 mL/min, followed by a wash of 100 % A for 5 min. Mobile phase B was changed from 0 to 100 % from 20 to 22 min and held for column equilibration until 30 min.

Given the nature of LC-MS/MS, MRMs were used to measure metabolites of interest. To observe isotopologues of a given metabolite, whether it be glucose or downstream products of glucose metabolism, MRMs were set up to include the multiple daughter isotopomers of a given parent isotopologue. For example, alanine M+1 MRMs were set up to consider that an M+1 parent isotopolgue could yield an M0 or M+1 daughter ion, depending on the carbon lost in fragmentation. Alanine $M+1_{+1}$ and alanine $M+1_{+0}$ (i.e. an alanine with one labeled carbon in the parent compound, with one or zero carbons labeled in the daughter ion respectively) signals were then summed to yield a parent alanine M+1 isotopologue signal.

Data processing and normalization

Chromatograms were processed using TargetLynx under MassLynx version 4.1 and exported as ion counts to be processed in R (version 3.3). For each sample, technical duplicates

were injected, along with QC samples every 10 injections, which was comprised of a pool of all samples and used to correct for analytical drift. Before QC-based corrections, a ratio matrix was derived from the raw data. For every metabolite, the ratio of each isotopomer to the unlabeled (M0) metabolite was stored, such that natural abundance isotope correction could be performed after QC correction. For every metabolic feature, a linear regression function fit to the QC data was used as a normalization factor for the samples as a function of run order (Dunn et al., 2011). Technical replicates were then averaged, and each sample was normalized to the injected blue dye (MRMs for unique m/z's found in the dye were tuned for ESI+ and ESI- analyses, Table 4.1), after which daughter isotopomer values were then summed to yield single parent isotopologues (e.g. alanine $M+1_{+1}$ and alanine $M+1_{+0}$ to form alanine M+1) in both the processed data matrix and the ratio matrix. The processed matrix was then rescaled back to 'raw' isotopologue distributions through division by the ratio matrix, so that natural ^{13}C abundance isotope correction could be performed, based on previous methods (Moseley, 2010). Values were then converted to relative abundance by dividing an individual isotopologue value to the sum of all detected isotopologues for a given metabolite.

Rhythmicity analysis and statistics

PCA was performed using SIMCA-P v14 (Umetrics AB, Umeå, Sweden) as a visualization of successful QC-based data correction (i.e. tight clustering of QC samples around the origin in Principal Component space) and to remove any outlier observations residing outside the Hotellings' T^2. Supervised clustering was performed using OPLS-DA with 7-fold cross validation as performed in SIMCA-P.

Rhythmicity testing of isotopologue relative abundances or metabolite pool sizes was performed with JTK_CYCLE under the MetaCycle package in R (Hughes et al., 2010, Wu et al., 2016) to test for 12 or 24-hour period lengths. Given the biological replicates consisted of separate days of injection and sacrificed flies (as opposed to resampling the same flies), JTK_CYCLE was modified to run 500 permutations of the data to guard against day-to-day bias.

Of the four or five replicate days of injections, for each injection ZT, two replicates were randomly chosen to serve as duplicates, while two of the remaining data points were randomly selected as a second day of testing, yielding a 48-hour dataset for circadian wave-fitting. In this way, a distribution of adjusted p-values could be derived from JTK_CYCLE analysis to mitigate daily technical variance. The default cosine function was used to assess significance, reported as the median p-value of the 500 permutations.

To test for non-circadian time-dependent changes in relative abundances or pool sizes, univariate tests were performed in a pairwise manner. A Bartlett's test was first used to decide between parametric (Student's T-Test, $F>0.05$) or nonparametric (Wilcoxon text, $F<0.05$) tests of means. A single time point was either tested to all other time points within genotypes, or compared across genotypes at the equivalent time point. A Benjamini-Hochberg (Benjamini and Hochberg, 1995) correction was performed to account for multiple testing using the 'stats' package in R.

4.4 Platform Rationale

Aligning technical approach with biological objectives

The workflow for the *in vivo* isotope tracer platform is outlined in Figure 4.1. Each step in this process required careful consideration of separating biological variance from unwanted technical and analytical variance, all with the prevailing objective of understanding dynamic metabolic processes across circadian time. The injections serve to mimic a both pulse-chase experiment and acute feeding bout, through injection of an isotope tracer and collection of the flies after a fixed time interval. Feeding flies a tracer through their food source with respect to circadian time would require enforcing appreciable feeding at the time point of interest, most likely through a fasting paradigm, which may confound interpretation of the results. A feeding approach may still be useful to answer questions regarding control of fasting/refeeding responses by the circadian clock but is outside the scope of this project. In order to remove the influence of continued feeding, these flies were placed in vial without food during the glucose incubation

period. Of note, these flies were anesthetized on ice, rather the preferred method of carbon dioxide. Initial anesthetizations under carbon dioxide revealed increased lactate M+3 compared to ice, which most likely reflects increased glycolysis under hypoxic conditions. While cold exposure may also exhibit some confounding metabolic influences, hypoxic conditions likely alter physiology more quickly than ice exposure. The duration of tracer exposure is critical to understand the acute response to nutrients under circadian cycles. One hour of incubation with the labeled glucose was considered a tradeoff between sufficient sampling resolution for circadian analysis and obtaining appreciable isotopic enrichments for LC-MS detection. Given the genetic clock can respond quickly to environmental stimuli (Myers et al., 1996), shorter incubation times would be ideal, but come at a cost of fewer detectable isotopologues. In time, with improved metabolomics technologies, shorter incubation periods with higher time resolution can provide additional clarity to metabolite flux cycles.

Analytical strategy

LC-MS metabolomics has traditionally been bifurcated into untargeted analyses on a high-resolution mass spectrometer and targeted analyses on a quadrupole or ion-trap instrument (Cajka and Fiehn, 2016). This untargeted versus targeted chasm has been crossed with recent advancements in MS instrumentation, facilitating analysis of hundreds of small polar metabolites on quadrupole instruments with an MRM approach (Rhoades and Weljie, 2016, Yuan et al., 2012). The difficulties of metabolite identification are substantially alleviated with fragmentation information, leading to faster biological interpretations of these datasets. These analytical advancements can also provide deeper analyses of isotope labeling patterns, but encounter some unique limitations under this current study design. Isotope enrichment analyses on GC-MS benefit from increased chromatographic separation, and thus only the parent isotopologue mass is required to uniquely identify metabolites. However, fragmentation is required for sufficient specificity of a given metabolite on LC-MS operating at unit mass resolution.

A glucose tracer with only one ^{13}C produces significantly fewer uniquely labeled products than from $^{13}C_6$ glucose. While fewer labeled carbons entails a simpler LC-MS analysis, singly labeled glucose will not probe metabolism as deeply as $^{13}C_6$ glucose. However, with fully labeled glucose, many more MRMs must be considered. For instance, to analyze serine M+2, MRMs must be incorporated to consider ^{12}C or ^{13}C carbons in the daughter ion. The possibilities scale considerably with an increasing number of carbons, thus requiring hundreds of MRMs for even tens of metabolites (Choi et al., 2012). During this method development process, MRMs were trimmed based on both realistic expectations of labeling patterns from $^{13}C_6$ glucose and trial and error. For example, one may expect to find AMP M+5 from $^{13}C_6$ glucose, but not M+10. Additionally, some isotope enrichments were undetectable in e.g. lysine from pilot labeling studies and were removed from the LC-MS analysis to improve S/N for other observable metabolites. However, future experiments which utilize other tracers will need to test for new enrichment possibilities, especially if one is to design these experiments as an “untargeted” isotopolomics approach.

Data processing and correction measures

Reducing unwanted variance in large-scale circadian metabolomics experiments is a nontrivial endeavor (Dunn et al., 2011). Preparation and run order should be randomized to reduce bias, and QC samples incorporated to account for analytical drift, which typically comprise of a pool of the biological samples or a standard mix. The QCs can be used to both correct for drift and remove spurious metabolites with highly response variance. This approach is common in metabolomics, including large circadian metabolomic experiments with high sampling resolution (Krishnaiah et al., 2017). These procedures can be integrated into large-scale isotope enrichment analyses but must be tailored to address unique challenges. Analytical drift correction requires fitting regression curves to the quality control samples, after which correction factors are imputed and applied to the non-QC samples. Since this approach is performed for each metabolic feature individually, every variable will then be normalized to a scale that centers around 1. For

downstream multivariate statistical analysis, this approach is acceptable, however isotope enrichments must account for naturally occurring ^{13}C and undergo natural abundance correction (Moseley, 2010). If every isotopologue is treated as a separate metabolic feature and corrected to its own QC values, the scale of the isotopologue signal to the unlabeled metabolite is lost. One method to correct for natural abundance would be to use the ratio of the raw ion counts of a given isotopomer to the unlabeled M0 signal as a rescaling factor after QC correction. Once isotopomers are summed to form the isotopologue data matrix, natural abundance correction could be performed to arrive at a final normalized concentration.

An obvious criticism of this platform is the intentional disregard of the fragment labeling information embedded in the isotopomer MRM analysis. Positional labeling information is significantly more specific to reaction activities than parent isotopologues, however can only be obtained from NMR analysis or extensive MS/MS methods, for example the 47 tandem mass isotopomers which were used to obtain complete isotopomer distribution analysis in aspartate (Choi et al. 2012). Prodigious additional work will be required to extend the labeling platform described here and translate fragment labeling formation into significant biological inference.

Relative abundance as unique variables for periodic and multivariate statistical analysis

Under this platform, every isotopologue is treated as a unique variable for statistical analysis. However, the endogenous pool size of a given metabolite likely varies across both circadian time and experimental conditions. We observed that when plotting the corrected isotopologues against their M0 signals, the values generally moved in tandem, implying that metabolite pool size changes drove the variance in isotopologue signals. Additional scaling was necessary to account for the pool size fluctuation and treat each isotopologue as a unique variable. Each isotopologue was converted to a relative abundance metric by dividing the isotopologue signal by the sum of all the given metabolite's isotopologues (Buescher et al., 2015). This approach seemed reasonable to gauge qualitative flux comparisons across time and genotypes, however may suffer from some inaccuracy if full isotopologue distributions are not

detectable. For example, while it is unreasonable to expect large AMP M+10 signals, if appreciate labeling collectively occurs in M+6 though M+10 isotopologues, and only AMP M0 through M+5 is analyzed, the relative abundance calculation will not completely account for pool size, but is still likely to be reasonably accurate for this particular experimental design.

After extracting relative abundances, each isotopologue can be tested for significant rhythmicity. A unique design feature in this *Drosophila* injection platform is the replicated sampling across multiple days. Rather than a continuous collection interval, which is common in cell or mouse experiments, replicate injections were performed on separate days, as five replicates on a single day of injections is infeasible (15 flies were injected at each time point for each replicate, which would take roughly 15 minutes). To guard against false positives and day-to-day variance, replicates were randomly permuted, while maintaining the circadian time order, and subsequently retested for periodicity. This approach is not common practice in the circadian field, but should be considered given the same animal is not typically resampled continuously and is instead sacrificed as a replicate data point. In addition, these statistical algorithms benefit from replicates across multiple days of testing, thus replicates data points were sampled to produce a time-series of two replicates across two days.

Metabolomic data analysis benefits from dimensionality reduction approaches such as PCA and OPLS-DA (Kotlowska, 2014). PCA is particularly useful to assess the quality of data normalization measures and outlier detection, and served similar functions in this high-dimensional isotopolomic dataset. Additionally, OPLS-DA has not been used in isotope analysis, but is particularly useful here to compare isotope enrichment profiles across genotypes and circadian time. As a dimensionality reduction technique, OPLS-DA can also guard against the influence of collinear X variables, which is inherent in the analysis of relative abundances. Given the influence of feeding and light on behavior and metabolism, isotope patterns which exhibit diurnal but noncircadian patterns would be otherwise unnoticed without these additional analyses.

4.5 Results

Glucose injection does not alter activity patterns

Supraphysiological doses of glucose may have unintended consequences on behavior and activity, thus we sought to inject a glucose bolus which would strike a balance between analytical detection and normal physiological response. After testing a range of concentrations (up to 3M), 1M glucose injections were deemed an adequate tradeoff of detectable isotope enrichment and animal safety, as flies empirically recovered at roughly the same rate from anesthetization as non-injected controls. Post-injection climbing assays did not reveal any significant changes between sham, PBS, and glucose injections in WT flies at ZT0 nor ZT12 (Figure 4.2). The only observable changes in climbing ability occurred across genotypes, at ZT0 between WT and *fumin* flies for PBS injections ($p=0.003$), and at ZT12 for PBS and glucose injections ($p=0.027$ and $p=0.014$ respectively). The locomotion activity of the WT flies decreased within the first half hour time bin for both PBS and glucose injections at ZT0 ($p<0.001$) and for glucose injections at ZT12 compared to the sham injection ($p=0.001$), however was not significant between PBS and glucose injections (Figure 4.3). All comparisons were additionally insignificant in the *fumin* mutants. These locomotion assays revealed an expected consequence of the injection injury when compared to a sham injection, however glucose did not impart any significant changes compared to the PBS injection control, demonstrating an appropriate dose for our experimental objectives. The injections were well tolerated, as less than one percent of all injected flies in the circadian experiments were killed.

Amino acid metabolism demonstrates significant rhythmicity is and perturbed with sleep loss

Serine M+3 relative abundance displayed significant circadian rhythmicity in WT flies ($p=0.02$) with a 24-hour period and a peak at ZT4, which was lost in *fumin* mutants ($p=1$, Figure 4.4A). Additionally, the relative abundance was significantly higher at ZT4 in WT over *fumin* flies ($p=0.005$), reflecting increased morning flux into serine from glucose carbons. The serine pool

size was also significantly circadian (p=0.01, Figure 4.4B) in WT flies and dampened in *fumin* flies. Of note, the serine pool size did not significantly change between ZT0, ZT4, and ZT8, supporting the notion of increased carbon flux into serine amidst a stable pool size during the active phase. Glutamine M+2 displayed a significant 12-hour period in WT flies (p=0.03) and a 24-hour circadian cycle in *fumin* mutants (p=0.004, Figure 4.5A), however no significant genotype comparisons could be made as peak phases were in the morning hours for both genotypes. Interestingly, the glutamine pool demonstrated a 24-hour period in WT (p=0.001) but tested insignificant in *fumin* (Figure 4.5B). The glutamine pool size in WT flies did not increase from ZT0 to ZT4, unlike glutamine M+2 relative abundance (p=0.01), suggesting increased flux into glutamine carbons under steady-state metabolite levels.

Glutathione M+2 exhibits ultradian patterns in WT that are lost in short-sleep mutants

GSH M+2, a product of glutathione biosynthesis from amino acid precursors, exhibited a clear ultradian pattern in WT flies but not *fumin* mutants (Figure 4.6A). Since 4-hour sampling resolution is inadequate to test 8-hour ultradian periods, pairwise time comparisons were made within WT samples, revealing significant relative abundance changes between ZT4-8, ZT8-12, and ZT12-16 ($p<0.02$). GSH pool sizes yielded a similar time course, however dropped significantly at ZT12, as ZT0-12, ZT4-12, and ZT8-12 pairwise comparisons were significant ($p<0.02$). We reason the source of the M+2 signal comes from glycine, as the WT glycine pool size also displayed significant changes between ZT0-12, ZT4-12, and ZT8-12 ($p<0.02$). The glycine M+2 signal was too weak and variable to yield statistically significant results, but also aligns temporally with GSH M+2 relative abundance (Figure 4.6B).

Isotopolome yields unique profile in the early active phase

Univariate pairwise comparisons across genotype revealed 40 significant isotopologues ($p<0.05$, $q<0.2$, Table 4.2), with 38 occurring at ZT4. Unique metabolites in this list consisted of AMP, erythrose-4-phosphate, malate, glucose-6-phosphate, citrate, GSH, glutamate, ribose-5-

phosphate, glucose, serine, alanine, asparagine, fructose-1,6-bisphosphate, succinate, leucine/isoleucine, ornithine, proline, and lactate. Although a significant model could not be fit for discriminant analysis of all time points as unique Y variables, a scores plot from an OPLS-DA model to the WT samples visually confirmed a unique isotopolome at ZT4 (Figure 4.8A). Additionally, discriminant analysis of ZT4 observations to all other time points in the WT samples yielded a significant model (R^2X=0.175, R^2Y=0.938, Q^2=0.426, p=0.016), as well as discriminant analysis by genotype (R^2X=0.178, R^2Y=0.809, Q^2=0.51, p=1.29e-06, Figure 4.8B). As this LC-MS platform predominantly features smaller molecules as products of carbohydrate oxidation, rather than anabolic fates to glycogen and lipids, this data suggests a unique fluxomic profile in oxidative pathways early in the active phase under LD and normal sleep/wake cycles.

4.6 Conclusions

The amalgamation of systems-level oscillations in transcripts, proteins, and metabolites presents a formidable degree of complexity in the circadian systems. Theoretical constructs which aim to decipher metabolic flux rhythms based on phase offsets of key metabolic enzymes have recently been published (Thurley et al., 2017), but do not experimentally measure metabolite fluxes. Although tracer studies have been performed in cell models of circadian disruption (Papagiannakopoulos et al., 2016), stable isotope metabolomics has yet to be performed in a model that incorporates both the environmental and genetic drivers of the clock. Herein a novel platform is established to inject stable isotope tracers into *Drosophila*, and subsequently analyzed using custom LC-MS methodology and data-processing procedures. The injections were well-tolerated, and yielded insignificant changes in activity between glucose and PBS injections, suggesting a reasonable dose to study circadian physiology.

Flux into amino acid metabolism from glucose carbons exhibited a significant circadian pattern in WT flies, with a higher morning flux compared to sleep-deprived flies. Serine M+3 represents a new circadian biometric variable and likely results from a branch point from glycolysis (Pizer, 1963). Previous studies have analyzed steady-state pools of serine across

circadian time (Dallmann et al., 2012), however could not state the source of the constituent carbons. Serine has recently garnered attention in cancer metabolic studies, as it sits at the nexus of lipid biosynthesis, one-carbon metabolism, amino acid biosynthesis, and redox defense (Mattaini et al., 2016). Interestingly, these major metabolic processes have all been implicated in circadian rhythms or sleep (Katewa et al., 2016, Krishnaiah et al., 2017, Rhoades et al., 2017, Inoue et al., 1995), suggesting serine production from carbohydrates may be an important preceding step under circadian control. Future studies should consider tracking the fate of serine carbons into downstream products to clarify a particular role for rhythmic serine flux. This analysis may further elucidate if cyclic serine labeling derives from a temporal offset of downstream branch points, or cyclic rediversion of upstream carbohydrate carbons. Additionally, novel tracer designs have recently been developed to define compartmentalized reactions in glycine/serine metabolism and may add additional interpretability to the results noted here (Lewis et al., 2014). Serine is currently of interest in cancer metabolism, serving as precursor for biomass production. While adult flies may not require the same degree of anabolism as cancer cells, rhythms in oxidative and reductive metabolism have been noted elsewhere (Peek et al., 2013, Katewa et al., 2016), and support a hypothesis that circadian clocks serve to temporally separate catabolic and anabolic processes to promote metabolic efficiency (Bass, 2012).

Serine M+3 lost rhythmicity in *fumin* flies, suggesting altered diurnal demands in amino acid metabolism. WT flies under LD and functional clocks can anticipate energetic demands and nutrient availability and may consequently shuttle the carbohydrates consumed in the morning into secondary pathways such as amino acid anabolism. With perturbed sleep/wake patterns, the *fumin* mutants may require additional energy from amino acid catabolism, thus breaking the possible roles of cyclic serine production for anabolic purposes. Isotopically-labeled amino acid injections would further elucidate cycles of amino acid catabolism and anabolism under sleep deprivation. Glutamine M+2 also marks a previously undiscovered ultradian output. Given 12-hour cycles are likely driven through LD entrainment, noted for steady-state metabolite levels in Chapter 3, glutamine M+2 may reflect to feeding and activity rhythms. Oxidative metabolism has

been previously shown oscillate with ultradian cycles in cells and liver *ex vivo* (Peek et al., 2013), thus glutamine M+2 could serve as an alternate readout of metabolic flux through central energy and amino acid metabolism. A circadian rhythm was noted for glutamine M+2 in *fumin* mutants. This finding may suggest an ultradian profile in central energy metabolism regulated by clocks, sleep/wake patterns, and feeding cycles, however in the absence of normal activity cycles, shifts towards a circadian profile aligned with an intact genetic clock.

Glutathione levels and glutathione biosynthetic enzymes were previously reported to exhibit a diurnal pattern in *Drosophila* heads (Beaver et al., 2012), but did not demonstrate periodicity in active glutathione biosynthesis. GSH M+2 displayed a stark ultradian pattern in WT flies, however the glutathione pool size demonstrated a similar temporal pattern, which may suggest a steadier flux of carbons into a transient steady-state level than would be suggested by M+2 relative abundance alone. Nonetheless, glutathione pool sizes did not alter significantly in the light phase, suggesting a degree of rhythmic flux. Although glycine M+2 enrichments were too low to accurately detect at every time point, the temporal overlap between glycine and glutathione supports the hypothesis that the glycine addition via glutathione synthetase represents the ultradian step in glutathione biosynthesis, as glutamate and cysteine isotopologues did not display a similar pattern. Of note, glycine M+2 did not temporally overlap with serine M+3. While glycine M+2 may still form from serine M+3, rhythmicity in substrate rediversion may be responsible for the temporal offset of these isotopologues. Interestingly, Beaver et al. reported rhythms in fly heads for glutamate cysteine ligase activity, and found no temporal pattern in glutathione synthetase, however head and body mechanisms of glutathione biosynthesis and redox defense may differ. Additionally, GSH M+2 cycles were lost in *fumin* mutants, which likely stems from altered ROS production under sleep deprivation, perhaps consistent with the loss of ROS cycling noted in clock mutants (Krishnan et al., 2008).

A global isotopolome shift at ZT4 in WT flies suggests increased carbohydrate metabolism, which may stem from increased feeding and activity after lights-on. Given many of the metabolites detected in this method comprise small compounds instead of larger anabolic

products, these results may reflect a role for glucose in a morning ‘catabolic rush’. Anabolic products, including glycogen, lipids, and proteins, would likely be formed towards the onset of dusk in anticipation of fasting and decreased activity, as has been previously noted (Tsai et al., 2010). Additional analytical methods would be required to analyze anabolic fates for glucose carbons. Alternatively, the results from Chapter 3 highlight the importance of light-induced metabolome profiles, which may be mirrored here in a LD paradigm. Future experiments should consider isotopolome analysis under constant conditions to separate environmental and clock-driven flux profiles.

The stark difference in the ZT4 isotopolomes across genotypes may reflect altered metabolic demands after sleep deprivation. Locomotor activity in *fumin* mutants is considerably higher than WT at night rather than after lights-on, mirrored by high nighttime dopamine levels (Kume et al. 2005). Additionally, light has been shown to inhibit dopamine’s wake-promoting effects through upregulation of inhibitory dopamine receptors (Shang et al., 2011), further promoting the notion that altered glucose metabolism may stem from a response or compensation to nighttime hyperactivity. Additional models of sleep perturbation will clarify the relative importance of high dopamine levels versus alternative means of sleep deprivation.

This platform has unearthed a new dimension of circadian clock outputs, however may be confounded by some aspects of the current experimental design. Considering the diurnal activation of stress pathways (Spencer et al., 1998), the cyclic metabolic outputs noted here may be influenced by stress responses to the injection injury. Additionally, this experiment subjected flies to cold, followed by a light exposure necessary to perform the injections. These additional stressors and cues may not only influence metabolism, but reset the clock, mitigating the effect size on diurnal flux. The clock itself may also gate the degree of response to environmental signals, which has been observed by differential photic induction of gene expression in subjective morning versus night (Sutin and Kilduff, 1992). One hour of labeling may be sufficiently long to alter the clock phase after a brief light exposure, however likely did not completely reset the clock given the significantly circadian isotopologues observed here. The hour incubation period

ultimately stemmed from a tradeoff of isotope enrichment with acute flux analysis. Many of the isotopologues included in the LC-MS analysis yielded relatively high variance and low S/N, thus additional advancements in instrumentation would be required to shorten the label incubation period to alleviate concerns of stress and light exposure while maintaining broad coverage of the isotopolome.

While unprecedented in scope, this current analysis cannot detect particular isotopologues of interest. Given glycine M+2 does not temporally align with serine M+3, other downstream metabolites from serine carbons would likely also oscillate with a 24-hour period, including one-carbon metabolites and lipid synthesis products such as sphingosines and phosphatidylserines. Substantially more sensitive methodology will be needed to detect these isotopologues, as the serine M+3 relative abundance in this study never exceeded 12%. MRMs for GSH only considered M0 through M+5, rather than the full isotopologue distribution, in an effort to increase MS dwell times and decrease analytical variance in other isotopologues. Dedicated methods which account for all possible MRMs of glutathione may further elucidate the source of glutathione labeling patterns and corroborate previous enzyme expression data.

The open-circulatory system in *Drosophila* provides a unique advantage in the interpretation of this study design. While mammalian physiology is highly compartmentalized, the bathing of the fly's organs in the hemolymph likely leads to a more rapid mixing of tracer and subsequent exposure to the periphery, and may be amenable to single-compartment kinetic models in future study designs (Wolfe and Chinkes, 2005). In mammals, glucose labeling patterns may be largely driven by glucose uptake, rather than intracellular reaction rates. In this *Drosophila* paradigm, this concern is mitigated by an assumed mixing of glucose through the hemolymph, however rhythms in glucose uptake would need to be validated through tissue-specific enrichment analysis. The fat body is the largest and most metabolically active organ in the fly, and is likely responsible for much of the observed labeling patterns. However, fat body transcripts do not necessarily align temporally with whole-body or head transcripts (Keegan et al., 2007, Xu et al., 2011), which will likely decrease the number of observable circadian

isotopologues from body analysis. A combination of new LC-MS methodology and fly preparation and extraction protocols will be required to discern tissue-specific labeling patterns, as many isotopologues noted even in body analysis displayed low S/N. Additionally, the metabolic source of a given isotopologue is poorly understood. While serine M+3 likely derives from glycolysis branching, metabolic paths leading to glutamine M+2 are less clear, and realistically may have many sources when one considers the complex network-structure of metabolism (Chokkathukalam et al., 2014). Additional bioinformatic approaches which track carbon flow throughout genome-scale metabolic networks will be required to understand the possible metabolic paths which lead to these experimental results, and is the subject of Chapter 5. These uncertainties may also be alleviated through additional MS/MS platforms which effectively utilize fragmentation labeling patterns. Ultimately, our current understanding of biochemistry has not kept pace with the advances in systems-level technologies, particularly in secondary metabolism (Hackett et al., 2016). Regardless, this acute labeling platform opens new doors for understanding circadian, sleep, and nutrient processing mechanisms. Apropos, a recent interest in the benefits of time-restricted feeding vis-à-vis the circadian clock raises questions regarding the temporal distribution of nutrient processing (Longo and Panda, 2016). Experimental designs which consider feeding paradigms, alternative isotope tracers, and clock disruptions have pragmatic applications in nutrition, therapeutics, and social determinants of health, and may improve the management of altered feeding and sleep schedules common in e.g. shift work.

Table 4.1 MRM transitions for injection dye used in signal normalization

Compound	**Parent Mass**	**Daughter Mass(es)**
Dye1 (ESI+)	453.00	217.00, 326.00
Dye2 (ESI+)	579.00	288.00
Dye3 (ESI-)	225.00	200.00
Dye4 (ESI-)	373.20	170.00, 333.00

Table 4.2 Significant isotopologues (p<0.05, q<0.2) across genotype

Isotopologue	**Time Comparison**	**P-Value**	**Q-Value**
AMP M0	4hr	4.74E-06	0.000783
AMP M+1	4hr	1.35E-05	0.001113
Erythrose-4-Phosphate M0	12hr	4.97E-05	0.008204
Malate M+1	20hr	5.42E-05	0.008939
Glucose-6-Phosphate M+6	4hr	0.000503	0.027657
Glucose-6-Phosphate M0	4hr	0.000976	0.040243
Citrate M+1	4hr	0.001445	0.047686
GSH M0	4hr	0.002447	0.062446
Ribose-5-Phosphate M0	4hr	0.0029	0.062446
Hexose M+6	4hr	0.003028	0.062446
AMP M+2	4hr	0.003663	0.067158
Serine M+3	4hr	0.005092	0.08402
Asparagine M+1	4hr	0.005982	0.08973
Hexose M0	4hr	0.006779	0.093208
Alanine M+3	4hr	0.009335	0.118477
Ribose-5-Phosphate M+4	4hr	0.010308	0.121487
Asparagine M0	4hr	0.013238	0.139188
Asparagine M+4	4hr	0.014246	0.139188
AMP M+5	4hr	0.014341	0.139188
Serine M+2	4hr	0.016467	0.150951
Glutamate M+4	4hr	0.018159	0.157693
Fructose-1,6-Bisphosphate M+3	4hr	0.021071	0.163972
GSH M+5	4hr	0.021837	0.163972
Alanine M0	4hr	0.021863	0.163972
Hexose M+3	4hr	0.02347	0.164591
Glucose-6-Phosphate M+3	4hr	0.02394	0.164591

Succinate M+2	4hr	0.025768	0.168292
Leucine/Isoleucine M+5	4hr	0.026519	0.168292
Asparagine M+2	4hr	0.028571	0.168367
Glucose-6-Phosphate M+5	4hr	0.028571	0.168367
AMP M+3	4hr	0.029997	0.169458
Ornithine M+1	4hr	0.031106	0.169458
Ornithine M0	4hr	0.031838	0.169458
Proline M+4	4hr	0.037125	0.187679
Proline M+5	4hr	0.038418	0.187679
Lactate M+1	4hr	0.038673	0.187679
Lactate M+3	4hr	0.04137	0.194402
Serine M0	4hr	0.044428	0.194402
Hexose M+5	4hr	0.044693	0.194402
Succinate M+4	4hr	0.044771	0.194402

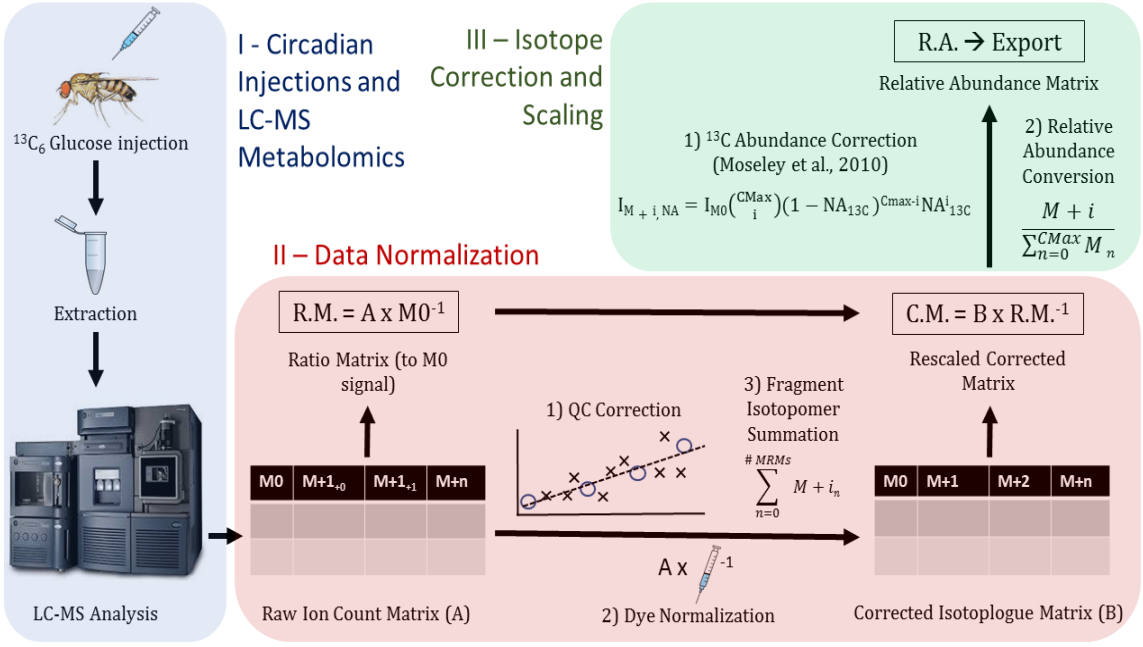


Figure 4.1 Workflow for *in vivo* circadian isotopolomics platform. After *Drosophila* injections and LC-MS analysis, data is normalized to QC samples and injection dye signal, followed by isotopomer summation to yield the isotopologue matrix. These values are then scaled according to ratios of individual isotopologue ion counts to the unlabeled M0 metabolite before natural ^{13}C abundance correction. Data is exported as relative abundance to the total metabolite signal.

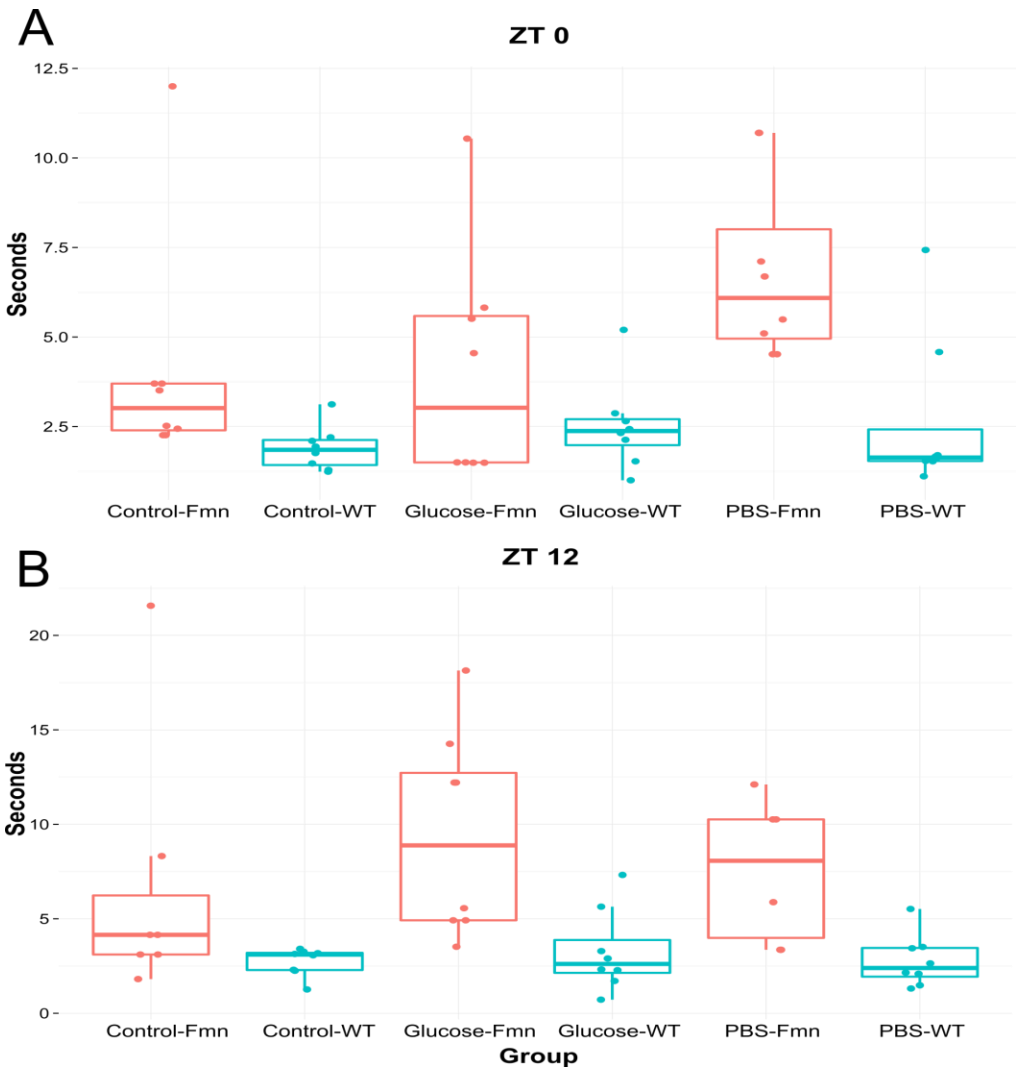


Figure 4.2 Geotaxis assay at ZT0 (A) and ZT12 (B), measured in time to climb 4cm one hour after injections (n=5-8). Boxplots are colored by genotype (Fmn – *fumin*; WT – wild-type). No significant changes were noted within genotype at either time point.

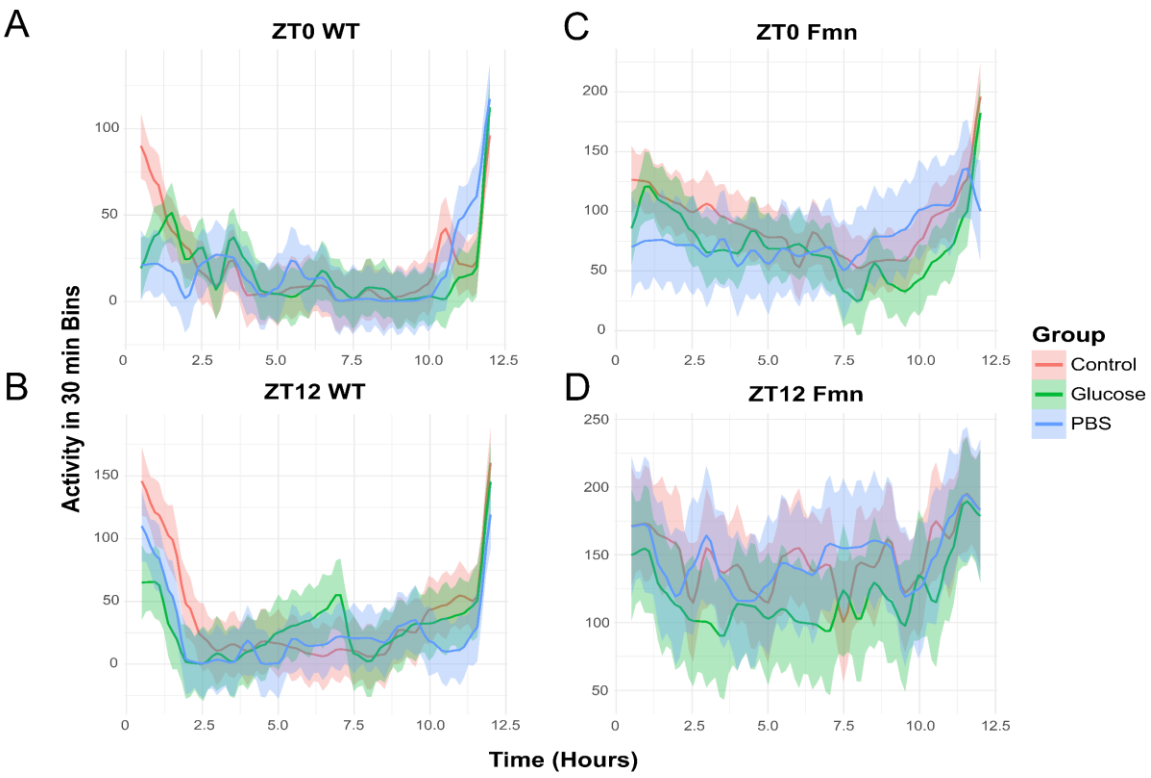


Figure 4.3 Locomotion assay by genotype and time. Glucose and PBS injections decreased activity compared to sham injection in WT at ZT0 (A, $p<0.001$), and for glucose at ZT12 (B, $p=0.001$), while no differences were noted in *fumin* mutants at either time point (C, D).

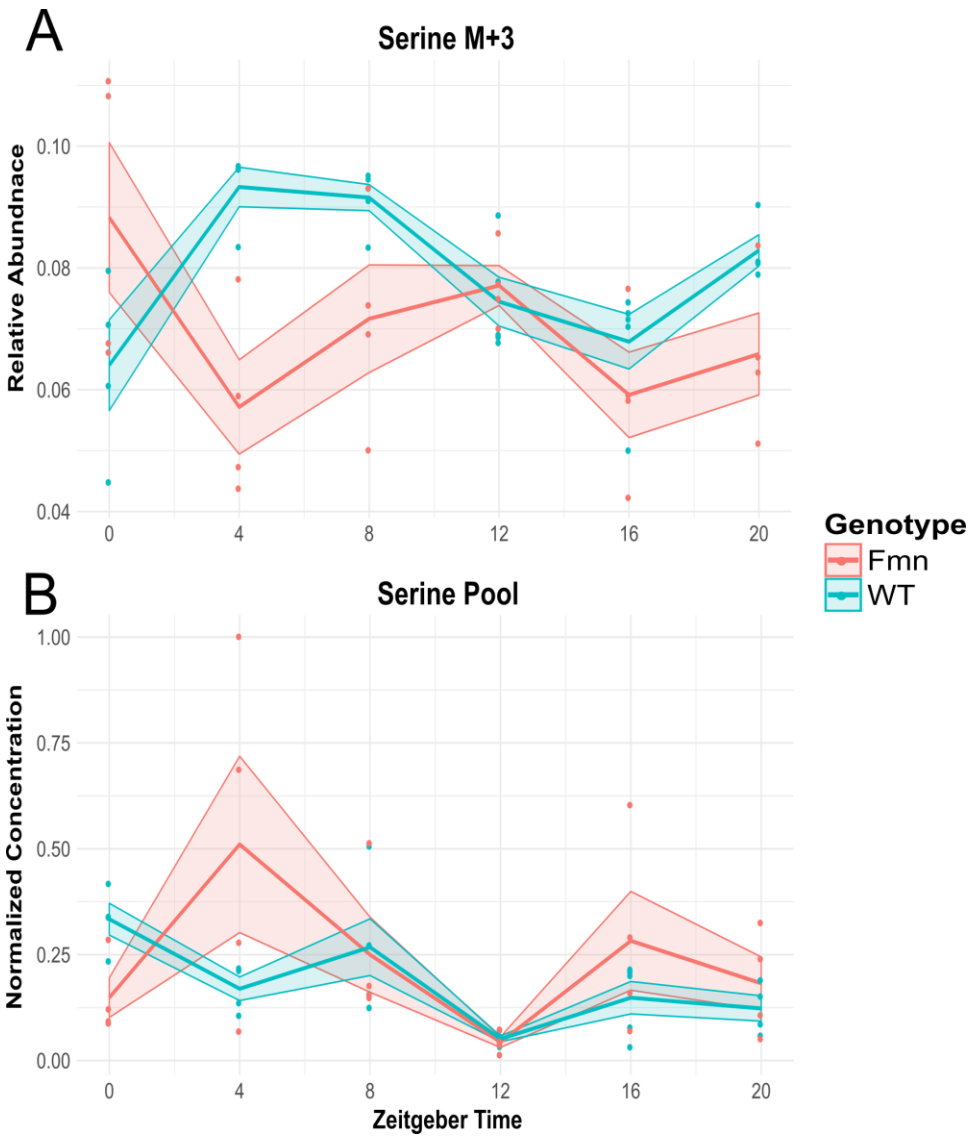


Figure 4.4 Relative abundance for serine M+3 (A) and normalized concentration of serine pool size (B) by genotype across time. Serine M+3 exhibited a significant 24-hour period in WT through JTK_CYCLE analysis (p=0.02) and significantly increased over *fumin* (Fmn) flies at ZT4 (p=0.005). Serine pool size did not significantly change between ZT0 and ZT8 in WT, demonstrating an increased flux of glucose carbons into serine.

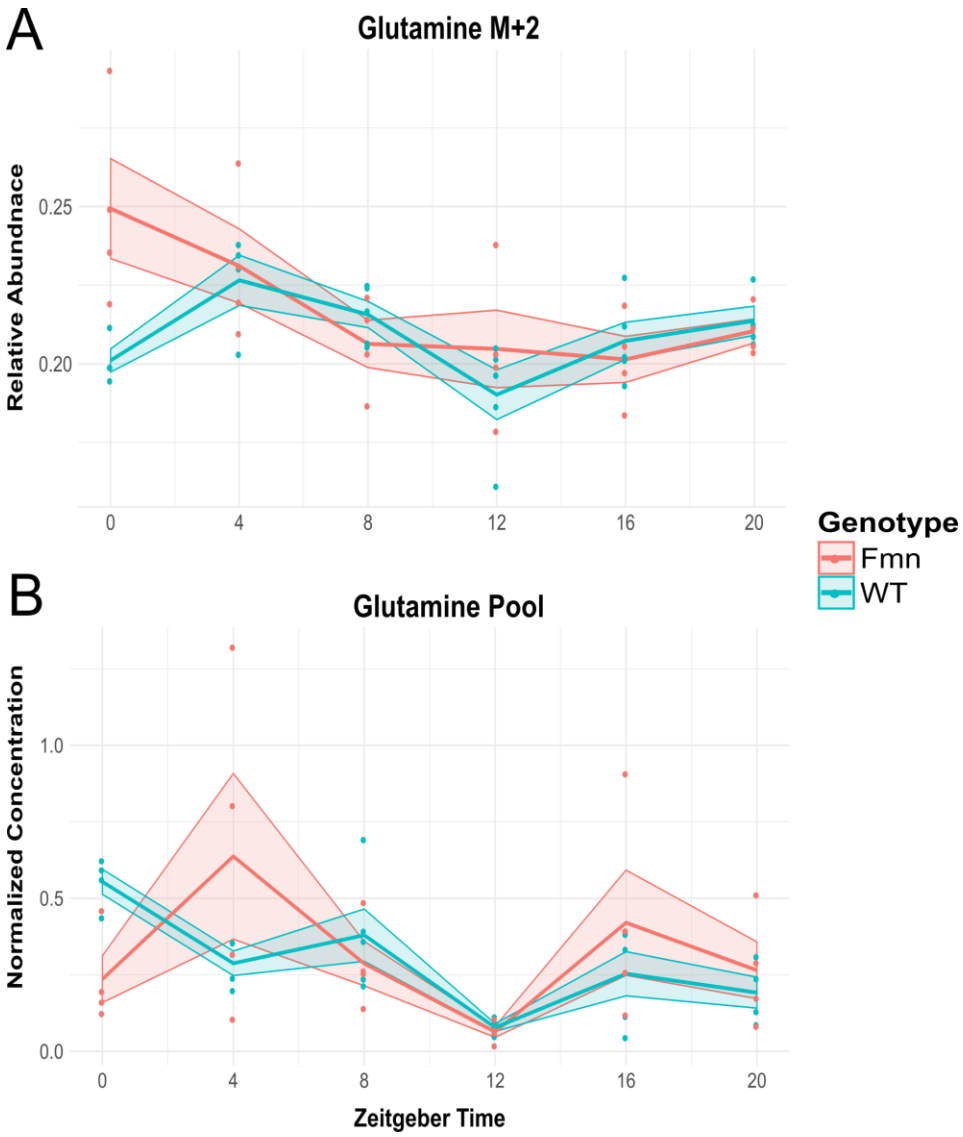


Figure 4.5 Relative abundance for glutamine M+2 (A) and normalized concentration of glutamine pool size (B) by genotype across time. Glutamine M+2 exhibited a significant 12-hour period in WT (p=0.03) and 24-hour period in *fumin* (Fmn) flies (p=0.004) through JTK_CYCLE analysis. Glutamine pool sizes displayed a significant 24-hour period in WT (p=0.001).

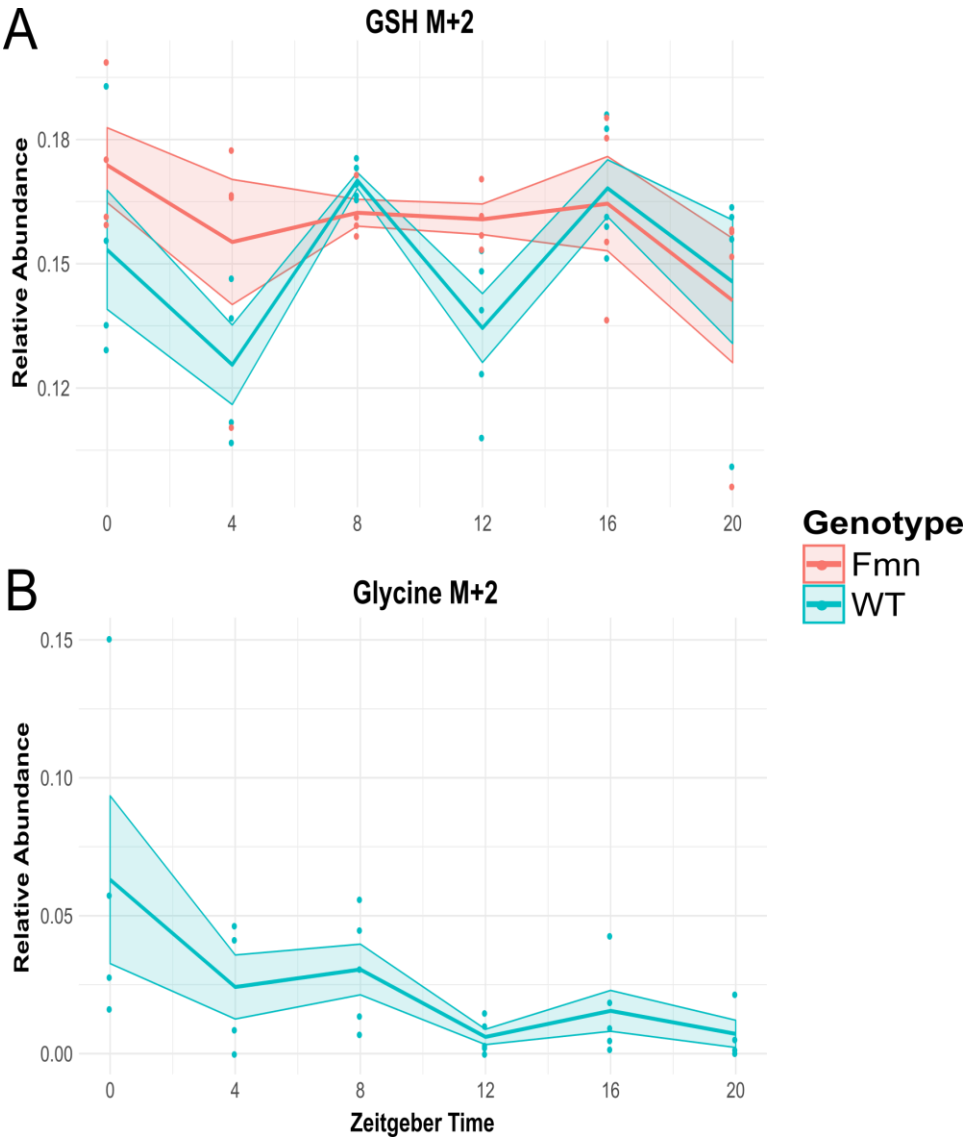


Figure 4.6 Relative abundance for GSH M+2 (A) and glycine M+2 (B) by genotype across time. GSH M+2 relative abundances significantly altered for WT in pairwise comparisons of ZT4-8, ZT8-12, and ZT12-16 ($p<0.02$), while relative abundances did not significantly change at any pairwise timepoint comparison for *fumin* flies. Glycine M+2 did not yield any significant changes, however displays an ultradian pattern which overlaps with GSH M+2 and suggests glycine as the carbon source for flux into GSH biosynthesis.

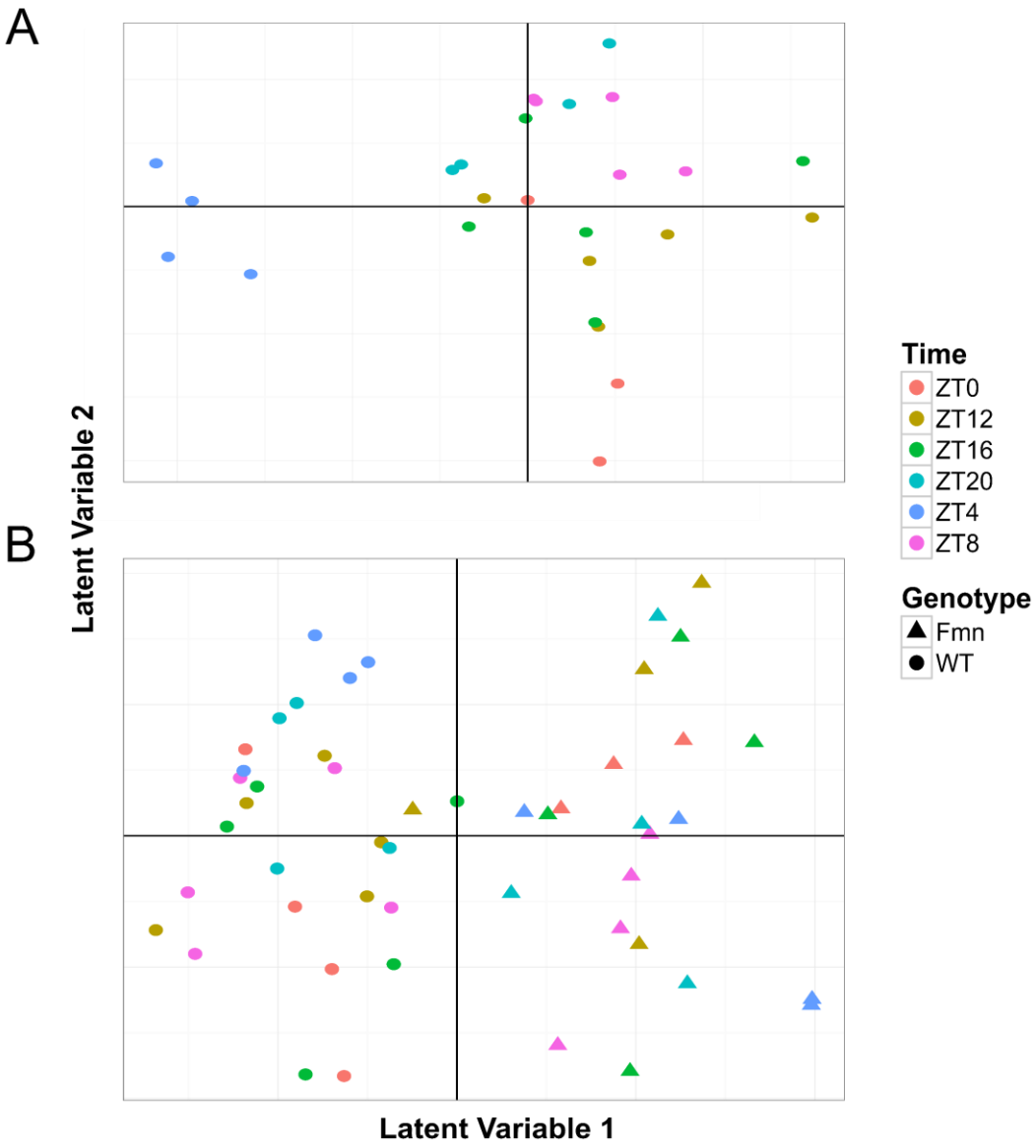


Figure 4.7 OPLS-DA scores plots. All time points were modeled as Y variables within WT samples to demonstrate unique ZT4 isotopolomic profile (A, ns), and a pairwise model by genotype, with all timepoints considered (B, R^2X=0.178, R^2Y=0.809, Q^2=0.51, p=1.29e-06).

CHAPTER 5 – *In silico* isotopologue-mapping to trace metabolic pathways at genome-scale

5.1 Abstract

Isotope tracers have greatly enhanced metabolic research, however as metabolomic analyses probes deeper into secondary metabolism, the interpretability of increasingly complex isotope enrichment patterns remains limited. Herein is described a novel computational approach, termed the IsoPathFinder, which considers all possible paths from the input isotope tracer to the isotopologue of interest through atom-resolved reaction networks at a genome scale. This algorithm is flexible, capable of rebuilding tracer-based metabolic networks from any tracer of interest and subsequently searching for any user-specified isotopologue. As a proof of concept, this approach uncovered possible routes to rhythmic isotopologues observed in Chapter 4. Canonical serine biosynthesis paths were calculated alongside two thermodynamically-favorable alternative paths through pyruvate M+3, which utilize an oscillating pyruvate kinase enzyme and confirm a glycolytic contribution to serine flux. 7246 paths were discovered for glutamine M+2, thus requiring a trimming procedure based on physiological relevance, expression datasets, and thermodynamics, to yield a subset of the eight likeliest paths. Most of these paths contained the TCA cycle, which confirms suspected cataplerotic feeding of the TCA cycle into amino acid pools. The shortest paths to GSH M+2 derived from glycine M+2, as expected. Additional routes via cysteine M+2 were discovered at longer path lengths, and may be driven from cyclic activity in glutamate cysteine ligase. This approach is presented as a proof of concept to confirm known biochemistry and propose new metabolic mechanisms, and opens doors for additional experimentation and metabolic discovery through simulation of alternative isotope tracers.

5.2 Introduction

The isotopologue data presented in Chapter 4 presents a set of novel findings regarding diurnal patterns in metabolic flux, generated from a bespoke technical and analytical metabolomics platform. On its own, this approach provides a clearer assessment of metabolic phenotypes over static measurements such as transcriptomics and steady-state metabolomics. However, we are far from fully understanding the enzymatic processes which produce observable isotopologues amidst the complex network of metabolism. Only a small fraction of the vast possibilities in metabolism are ever considered with current implementations of experimental designs and analytical assays, and subsequent interpretations on the scaffolds of restricted canonical biochemical pathways. While glucose tracing experiments have revealed glycolysis is in fact a major active hub of metabolic flux networks (Park et al., 2016), there remains vastly unexplored territory in alternative fates of isotope tracers. Bioinformatic tools have been recently developed to abstract labeling patterns from high-resolution MS metabolomics data (Huang et al., 2014, Capellades et al., 2016), however considerable work remains in metabolite identification and biological interpretation. Additionally, attributing labeling information to specific reaction activities is difficult without atomic detail, which would require NMR or extensive MS fragmentation (Choi et al., 2012). The rapid growth of reaction annotation and metabolite identification presents a unique opportunity in systems-level biomedical research to unbiasedly explore all enzymatic possibilities *in silico*. Only recently has computing power been commensurate to predict fluxes to scale (Ebrahim et al., 2013), however these computational platforms do not yet exhibit the granularity of atomic resolution. No tool currently exists to track carbon transfers for every reaction in large metabolic networks and subsequently reconstruct all possible metabolic routes which lead to observable isotopologues. These two objectives served as the motivating factors for development of a new computational approach for *in silico* metabolic reaction modeling described here.

The findings in rhythmic amino acid and glutathione metabolism highlighted in Chapter 4 would benefit from such an approach. For example, carbon flow into glutamine M+2, serine M+3,

and GSH M+2 may have many sources, and while glycolysis and TCA cycle are likely contributors to these isotopologues, other possibilities are yet unexplored. In addition to developing a novel approach to explore alternative routes, a means to predict the most likely routes is presented through the incorporation of previously published datasets (Xu et al., 2011, Gill et al., 2015, Tepper et al., 2013), This new platform has shed light on the immensely complex nature of metabolism, however can generate hypotheses with heightened probabilities of the metabolic paths which contribute to observable data. This algorithm is designed with the intention of flexibility to both build sets of labeling possibilities from any input tracer of interest and search for paths to any isotopologue of interest, and is presented here as a proof of principle to both corroborate prior biochemical knowledge and provide novel alternative paths to significantly cyclic isotopologues from Chapter 4.

5.3 Algorithm and Workflow

Materials

All *in silico* reaction modeling and path searches were developed within Jupyter notebooks (v4.2.3) in Python v3.5. Scripts were run under a Ubuntu 15.10 virtual machine on an Intel Xeon E5-2650 system (24 cores, 2.1GHz) with 64GB of memory. Files which include the simplified molecular-input line-entry system (SMILES) formatted atom-mapping reaction solutions, fly-specific metabolic models, and reaction links from MetaCyc 'RXN' format to enzyme commission (EC) numbers were obtained from MetaCyc (Caspi et al., 2016). Conversions of SMILES format to common names for metabolites were generated from the PubChemPy Python module (v1.0.3) using a custom script, which ultimately sources information from the PubChem database (Kim et al., 2016). Additional metabolite naming mechanisms were manually changed to user-defined names for ease of interpretation, for example 617-45-8 to Aspartic acid (listed by chemical abstracts service (CAS) as the first synonym entry for the SMILES conversion of Aspartic acid in PubChem). Additional SMILES atom-mapping reactions were manually added or curated for completeness of central energy metabolism, including glycolytic and TCA cycle

reactions. Results from the *in silico* reaction modeling were converted into queriable dictionaries and used to construct metabolic paths under the IsoPathFinder algorithm. The list of all possible labeling paths was trimmed using multiple data sources: significantly circadian enzymes from *Drosophila* transcript studies were derived from Xu et al., 2011 and Gill et al., 2015, and estimated Gibbs free energies for each enzyme were based on calculations from Tepper et al., 2013.

Enumerating all possible isotope transfers

The general workflow for isotope reaction modeling, followed by reconstruction of metabolic routes, is outlined in Figure 5.1. Reactions which contain atom-mapping solutions of metabolites in SMILES formats, from the MetaCyc database of annotated *Drosophila* enzymes, were used to compile the starting metabolic network. Atom-mapping solutions consist of annotated reactions where atoms are numbered for all reactants and matched to atoms in all products, for example the 1' carbon of glucose to the 1' carbon of glucose-6-phosphate from the hexokinase reaction. Some enzyme and metabolite annotations were manually curated for accuracy and interpretability, including glycolysis and TCA cycle. While the objectives here included holistic and unbiased reaction modeling, given the $^{13}C_6$ glucose tracer employed in Chapter 4, any metabolite routes which fan out from central carbon metabolism would not be accurate without thorough curation of these core reactions. All possible carbon transfers were executed in an iterative round-by-round fashion, where the user specifies the labeled metabolite of interest (e.g. $^{13}C_6$ glucose), followed by a trimming procedure of the model to produce the starting set of reactions which contains the tracer. As the input tracer is defined by a SMILES string, the user may additionally specify positional labeling, such as 1,2-^{13}C glucose. The carbons from the tracer are then matched to the carbons in the products of the starting reaction set, and subsequently 'labeled'. These simulations are then stored as a Python-formatted dictionary, where the keys for the first round of reactions are the newly labeled products, and the values contain both the labeled reactants and the enzyme(s) which catalyzed the reaction (Figure 5.2).

These results can then be queried in the IsoPathFinder algorithm to build metabolic routes which originate from the input tracer and are exported as a .csv file (Figure 5.2C). The list of newly labeled products is then used to seed the second round of reactions. As in the first round, all reactions which contain any of these newly labeled compounds are selected from the network, followed by mapping all labeled carbons onto new products. In this way, the process can be repeated until the desired number of reactions is calculated. Initially, 30 rounds of reactions were deemed sufficient, especially given the possibility of glycolysis and multiple cycles through the TCA cycle. While 30 rounds may produce experimentally observable labeling patterns, the possible routes calculated in the IsoPathFinder algorithm quickly became computationally intractable, for example the more than 100,000 path possibilities for glutamine M+2 after 15 reactions. Consequently, all 30 rounds were never considered, as extensive code optimization and removal of many unrealistic reactions were necessary to even reach path lengths of 15. For instance, many reactions contained carbon dioxide, which is unlikely to appreciably contribute to observable labeling patterns in Chapter 4, and was thus excluded from the list of labeled compounds. Additionally, labeled compounds which did not have a match of SMILES format to a common name in the PubChem database were dropped. Future implementations of this approach may benefit from simulating reactions which contain multiple labeled reactants, however would present both a more difficult computational problem to solve and a substantially greater set of possible labeling reactions to interpret.

IsoPathFinder algorithm

After simulating all possible isotope transfers and rebuilding a new isotopically-labeled metabolic network, routes are built working backwards from the user-defined isotopologue of interest to the initial input tracer in round one. For example, if serine M+3 were found in round ten, all reactants which mapped to serine M+3 in round nine would be returned, followed by a search for all reactants which led to those product metabolites in round eight, etcetera. These routes were then appended to a repeated construction of paths for serine M+3 starting in round

nine, followed by starting from each of the preceding rounds until serine M+3 was not produced (for example, no reactions catalyze glucose M+6 to serine M+3 in round one). Currently this algorithm is capable of building paths for a maximum length of 15, however shorter paths can be defined by the user. Code optimization and/or computing clusters will be required to run longer path searches, as path searches of length 15 could take hours on a single computer. Some routes contained metabolites repeatedly across multiple reactions, which were removed to both further trim the number of possibilities and remove futile cycling. As a broader objective to achieve widespread usage, this program is designed to take a user input and build paths on any commonly observed isotopologue of interest, whether it be derived from experimental data or for purely *in silico* experimentation.

Reaction trimming

Given the number of possibilities, unlikely metabolite routes were trimmed using additional rationale and prior datasets. Some PubChem identifiers did not match to common names or other databases, such as metabolite 'AC1N7ZT6'. Paths which contained these IDs were manually trimmed if no alternative naming mechanism could be found. Additional steps were suited to meet the experimental objectives of deriving significantly rhythmic metabolic flux possibilities. Routes were trimmed which did not contain a circadian enzyme from two previous *Drosophila* transcriptomics studies (Xu et al., 2011, Gill et al., 2015). This approach may yield misleading results, as cyclic enzymes are not necessary to generate rhythms in metabolic flux (Thurley et al., 2017). However, this approach removed an appreciable number of paths and eased the interpretation of these results. Gibbs free energies were previously calculated for all annotated reactions (Tepper et al., 2013), based on an expanded formulation of the group contribution method (Mavrovounioitis, 1990). While these estimates may not be identical to *in vivo* enzyme kinetics for flies, these estimates were found to hold reasonably true when compared across other databases of enzyme thermodynamics (Goldberg et al., 2004). In addition, metabolic flux and enzyme efficiency are highly conserved processes under strong evolutionary pressure

(Park et al., 2016), which alleviates the concern of inaccurate free energies across species. These estimated Gibbs free energies were summed across all enzymes for each of the IsoPathFinder paths, and subsequently sorted by thermodynamic favorability. Future improvements to this algorithm may consider incorporating reaction irreversibility to further trim possibilities *a priori*. Unfortunately, even these steps cannot trim the possible routes to down to one, however by significantly reducing the possibilities and shifting probabilities towards a subset of paths, validation experiments can be designed to test relatively fewer hypotheses.

Separate Jupyter notebooks perform each aspect of this approach: enumeration of all possible isotope transfers from any input tracer of interest, path-building towards the isotopologue of interest, and reaction trimming based on the research objectives. These three scripts are available upon request.

5.4 Results

IsoPathFinder uncovers shortest set of paths from $^{13}C_6$ glucose to serine M+3 through glycolysis

The shortest path length to serine M+3 from fully labeled glucose consisted of nine reactions, with sixteen total unique paths. After trimming two paths which contain a carbon-carbon lyase of bacterial or plant origin (EC 4.1.1.39), eleven paths with a net positive ΔG, and one additional path without a circadian enzyme, two paths remained (Table 5.1). These paths included glycolysis, followed by conversion of pyruvate M+3 to serine M+3. Pyruvate kinase (EC 2.7.1.40), discovered to transcriptionally oscillate in fly bodies by Gill et al., was contained in both of these paths, thus providing predictions for the shortest thermodynamically favorable path to a rhythmic flux into serine carbons (Figure 5.3). Notably, the canonical serine biosynthesis pathway did not contain a cyclic enzyme, nor yielded a negative ΔG, and was removed from the initial set of paths. No unique paths of ten reactions were calculated with this algorithm, which may stem from a requirement of an additional intermediate before an eleventh reaction step. A path search of eleven reactions yielded 2262 unique paths. After trimming procedures, 61 paths remained.

Whereas the preceding metabolite in the four paths of length nine was pyruvate M+3, now 3-hydroxypyruvate, alanine, acetyl-serine, and cystathionine M+3 were alternative isotopologue sources. Interestingly, alanine M+3 demonstrates similar a similar temporal pattern to serine M+3 (Figure 5.4), providing additional evidence these paths may be relevant for the rhythmic flux noted in Chapter 4. However, isotopologues for hydroxypyruvate, acetyl-serine, and cystathionine were not included in this LC-MS analysis, but are discovered here from the IsoPathFinder algorithm as additional targets for future analyses. Most of these paths use glycolysis, however ten of these paths included the gluconeogenic enzyme phosphoenolpyruvate carboxykinase, catalyzing phosphoenolpyruvate M+3 to oxaloacetate M+3. Pyruvate M+3 was formed from oxaloacetate M+3 from the cyclic enzyme pyruvate carboxylase (EC 6.4.1.1).

Likeliest paths to glutamine M+2 derive from TCA cycle

Path lengths of fourteen steps were the shortest routes which yielded feasible routes to glutamine M+2 after trimming procedures, and produced 7426 paths. After trimming procedures, including removal of paths with non-physiological isotopologues such as ethylene and methyl-aspartate M+2, eight paths remained. For all paths, the preceding isotopologue to glutamine consisted of glutamate M+2, formed by alpha-ketoglutarate M+2. While glutamate M+2 did not significantly oscillate, the labeling pattern followed closely to glutamine M+2, corroborating these paths (Figure 5.5). The net ΔG in these paths ranged from -273kJ to -297kJ (Table 5.2), which were considerably more favorable than the highest value observed for serine M+3 paths (-68kJ), and contained upwards of seven cyclic enzymes. Half of these paths contained the TCA cycle, which included the most favorable path with a relatively high number of cyclic enzymes (glucosamine 6-phosphate synthase, pyruvate kinase, aconitase, isocitrate dehydrogenase, ornithine transaminase, and citrate synthase). The remaining paths did not include the TCA cycle, but instead formed alpha-ketoglutarate M+2 from 2-hydroxy-4-oxopentadioic acid M+2. However, given relatively fewer circadian enzymes in these remaining paths, we speculate the most likely paths to glutamine M+2 include the TCA cycle (Figure 5.6).

Glutathione biosynthesis likely an extension of rhythmic flux into serine M+3

A path search for GSH M+2 yielded 23 paths of eleven steps, which was trimmed down to two unique paths (Table 5.3). Each of these paths were a direct extension of the four shortest paths found for serine M+3, which now included a formation of glycine M+2 before GSH M+2 production via glutathione synthase. A path search up to twelve steps yielded an additional 24 paths after trimming, which all utilized glycolysis and contained glycine M+2 as the only preceding isotopologue (from serine M+3). Thirteen steps were necessary to find cysteine M+2 as an alternative source for GSH M+2 synthesis, which utilized a cyclic gamma-glutamylcysteine synthetase enzyme not found in glycine M+2 paths. GSH M+2 production from glycine were the shortest and most thermodynamically favorable paths, and represent the most likely paths to GSH M+2 (Figure 5.7).

5.5 Conclusions

The rapid expansion of stable isotope applications in recent years can greatly benefit from a commensurate tool which reconsiders all possible enzymatic processes now annotated in the post-genomic era. The approach outlined here represents the first attempt to map isotopologue data onto genome-scale metabolic networks at atomic resolution through a two-step bioinformatic approach. The first step calculates every carbon-transfer from an input tracer of interest in an iterative round-by-round manner. These simulations require a database with an extensive set of atom-mapping reaction solutions, which has only recently been developed and annotated (Caspi et al., 2016). Fortunately, while there exists room for improvement, MetaCyc represents a substantial effort by an active research community to compile such detail. To address the uncertainties described in Chapter 4, MetaCyc reactions were additionally curated around central energy metabolism to ensure proper tracking of glucose carbons, and reactions which did not appear likely to contribute to observable metabolic flux in *Drosophila* were removed, including bacterial enzymes. The microbiome connection to host physiology is of intense interest,

including flies (Shin et al., 2011), however given the bolus injection paradigm employed in Chapter 4, likely does not contribute appreciably to the observed isotope enrichments. Thus, with minimal additional work, existing databases can be effectively leveraged to explore new aspects of metabolism. Once all carbon transfers were simulated, both for all atom-mapping solutions from MetaCyc for general use and all fly-specific reactions for this study, this new scaffold of 'labeled' metabolism can be probed to discover all possible connections of the input tracer to the isotopologue of interest. This path-building step is designed to receive user input to calculate paths towards any commonly detected isotopologue, and can have broad applicability in isotope-based research. The IsoPathFinder produced many more possible paths through this new metabolic network than expected and ran into computational and interpretability constraints. The extent of this complexity is simultaneously exciting and humbling, but unfortunate given a metabolic path length of even 20 reactions seems physiologically feasible but cannot be achieved with this method. Additional algorithmic improvements will be necessary with a concurrent means to trim unlikely paths to alleviate the computational burden.

This algorithm uncovered a set of paths which trace carbons from glucose to serine through the shortest and most thermodynamically favorable paths. These paths utilize glycolysis and pyruvate kinase to form serine M+3 via serine deaminase (EC 4.3.1.17), however, some notable alternatives were trimmed. Serine biosynthesis is known to occur through 3-phosphoglycerate, catalyzed by phosphoglycerate dehydrogenase (Pizer, 1963). These paths were discovered with the IsoPathFinder, however were trimmed given a lack of circadian enzyme expression and a positive ΔG. Transcription may not necessarily correlate with reaction activities, and circadian enzyme expression may not be necessary for rhythms in metabolic flux (Thurley et al., 2017), especially considering the possibility that a cyclic pyruvate kinase enzyme may produce rhythmic flux upstream in glycolysis. An additional twelve paths were calculated without cyclic enzymes and thermodynamic trimming, which included the canonical serine biosynthetic pathway. Phosphoserine M+3 is the immediate precursor to serine in this pathway, but was not analyzed in the LC-MS platform described in Chapter 4. Additional experiments with expanded

LC-MS analyses may provide additional evidence for the relative sources of cyclic carbon to serine. Akin to metabolite pools, any given isotopologue is likely the product of multiple reactions, thus we can speculate that both these major paths may contribute to the findings in Chapter 4, especially given pyruvate M+3 did not yield a similar temporal pattern to serine M+3. Additional precursor isotopologues were discovered at paths of eleven reactions, including alanine M+3, which does resemble serine M+3 patterns. However, these paths may not provide significant additional biological insight, as the alanine M+3 is derived from pyruvate M+3 and would likely produce a similar metabolic phenotype to the shorter paths which do not contain alanine.

Gluconeogenic reactions were contained in some of these extended serine M+3 paths, which is not compatible with active glycolysis but may have interesting biological explanations. Organisms mobilize energy stores in anticipation of active states (McGinnis and Young, 2016), and likewise produce sugar via gluconeogenesis to further meet this energetic demand (Kida et al., 1980). The isotopologues representative of gluconeogenic flux, such as glucose-6-phosphate M+2, M+3, and M+4 (Wolfe and Chinkes, 2005), did not exhibit significant diurnal variance. However, with a supraphysiological glucose bolus, gluconeogenic flux may be overridden by glycolysis. Interestingly, glucose-6-phosphate M+2 and M+3 relative abundances were significantly increased during the light phase in WT compared to *fumin* flies ($p<0.03$), implying altered energetic demands in central energy metabolism. This computational approach thus provided evidence for cyclic serine flux through glycolysis, and revealed new alternatives to canonical biochemistry serine biosynthesis. Understanding both source and fate of serine carbons is important given its connection to numerous metabolic processes, and warrants additional experimentation through expanded LC-MS analyses and computational flux predictions (Mattaini et al., 2016).

Paths to glutamine M+2 which passed the trimming procedures were not observed until path lengths of fourteen reactions, and consequently produced numerous possibilities. These paths were rationally trimmed to predict glutamine production through the TCA cycle. Notably, these paths were very thermodynamically favorable compared to those for serine M+3 and GSH

M+2, and contained a series of circadian enzymes in the TCA cycle, which further increases the likelihood of these paths exhibiting cyclic activity *in vivo*. This approach also provided a link to overlapping isotopologue patterns from glutamate to glutamine M+2, which is particularly striking given 15 separate enzymes catalyze this last step in the glutamine M+2 paths. Alpha-ketoglutarate M+2 preceded glutamate M+2 in all these paths, however did not cycle. Alpha-ketoglutarate M+2 does exhibit significant but non-circadian diurnal changes in WT flies, thus raising the possibility of a cyclic contribution to glutamate and glutamine carbons through alternative mechanisms such as substrate rediversion. Paths which do not utilize the TCA cycle still utilize carbons through glycolysis and pyruvate, suggesting an important role for central carbon metabolism to downstream amino acid pools. Alternative paths uniquely contained 2-hydroxy-4-oxopentanedioic acid M+2, which was not detected on this LC-MS platform. As noted with serine M+3 paths, glutamine M+2 sources may not require cyclic enzymes, and alternative possibilities should not necessarily be omitted without additional experimentation. This procedure nonetheless trimmed a substantial number of possibilities, providing a more tractable set of hypotheses and shifting the probabilities towards more likely paths through the TCA cycle.

Steady-state glutamate pools robustly oscillate in flies (Chapter 3), with conserved rhythms in mice and humans (Hatori et al., 2012, Dallmann et al., 2012). Given the large pools of glutamine and glutamate, cyclic M+2 patterns may reflect a sink of carbons from upstream reactions in nitrogen and central energy metabolism. However, glucosamine-6-phosphate (GlcN6P) synthase, one enzyme which catalyzes the final step in glutamate M+2 to glutamine M+2, has been shown to oscillate in flies (Gill et al., 2015), and may reflect more active biological roles of amino acid flux. Hexosamine synthesis triggers protein modifications by N-acetylglucosamine (GlcNAc), which produces a circadian pattern of Bmal and Clock GlcNAcylation (Li et al., 2013). Manipulating GlcNAc levels can perturb both glucose homeostasis and circadian period length, perhaps serving as a metabolic sensor to coordinate with core clock functions (Kaasik et al., 2013, Kim et al., 2012). Additional experiments may consider the interaction of hexosamine and amino acid flux in relation to the core clock protein modifications.

Acetyl-CoA M+2 was featured in the TCA-dependent paths to glutamine M+2. Acetyl-CoA sits at the nexus of numerous intracellular processes, serving roles in both oxidative and reductive metabolism and epigenetic regulation as an acetyl source. As both oxidative metabolism and macromolecule biosynthesis have been shown to oscillate (Peek et al., 2013, Katewa et al., 2016), one may expect to find cyclic acetyl-CoA production. Additional LC-MS analyses would be required (Basu et al., 2011), as acetyl-CoA M+1 and M+2 were not detectable on the platform described in Chapter 4. Tracking acetyl-CoA carbon fates could shed light on the dynamic epigenome which has been previously observed through cycles in histone deacetylase 3 activity (Feng et al., 2011). A dependency on the clock would also be expected, as livers from *Clock* knockout mice show significantly altered acetylation profiles for enzymes in glycolysis, TCA cycle, and amino acid metabolism (Masri et al., 2013).

As expected, glutathione biosynthesis is most likely driven by glycine M+2 addition to glutamyl-cysteine via glutathione synthetase. This enzyme has not been found to oscillate in flies, however may still yield rhythmic glutathione synthesis through rhythmic substrate availability, as this last step in glutathione biosynthesis is rate-controlling. Glycine M+2 yielded a similar ultradian pattern to GSH M+2, however serine M+3 demonstrated a decoupled rhythm. As serine carbons feed other major metabolic pathways, temporal offsets in substrate rediversion likely produce the uniquely ultradian glycine and glutathione patterns. An extended path search revealed alternative paths to glutathione biosynthesis through cysteine M+2. While these paths were less thermodynamically favorable than glycine M+2 paths, they may be worth considering as glutamate cysteine ligase has been shown to not only oscillate transcriptionally, but also display diurnal patterns in enzyme activity (Beaver et al., 2012). Both glycine and cysteine exhibit relatively high variance on this LC-MS platform, which precludes cogent insights to glutathione carbon sources. Alternate metabolomic methods which accurately detect all glycine and cysteine isotopologues, and additional isotopologue products of serine metabolism, may further clarify the connection between amino acid and redox metabolism. These studies may additionally benefit from isotopically-labeled amino acid injections in *Drosophila*. The IsoPathFinder algorithm

ultimately provided supporting evidence to the expected source of glutathione carbons from glucose via glycolysis and amino acid metabolism.

While any of these speculations regarding functions of rhythmic flux into specific isotopologues may be appealing, substantial additional information will be required. As labeling patterns provide a clearer assessment of flux than transcript and protein levels, ideally data for every isotopologue along each path would provide a stronger basis for path selection, however is currently infeasible on this LC-MS methodology. For instance, given the extensive branching around acetyl-CoA, labeling patterns in histone acetylation and larger lipids would be required to gauge possible temporal offset in flux through other major branches of metabolism. There yet exists a large gap in enumerating paths to e.g. rhythmic glutamine M+2 production and understanding what role any or all individual isotopologues along those paths serve in chronometabolism. Consequently, we want to stress a conservative interpretation of these predictions from the IsoPathFinder algorithm, given the unexpected multitude of possible paths after merely fifteen reaction steps. Additional datasets may support the likelihood of a subset of these possibilities, however are prone to false negatives without considerably more detailed experiments with analytical detection methods commensurate to the number of variables embedded in these metabolic networks.

The significantly rhythmic isotopologues observed in Chapter 4 were processed in the IsoPathFinder algorithm to produce expected paths from glucose based on prior biochemical knowledge (Pizer, 1963, Buescher et al., 2015, Chaneton et al., 2012). Given the input glucose tracer, most paths utilized glycolysis, however the overabundance of glycolytic paths raises some concerns. During development of IsoPathFinder, glycolytic and TCA cycle atom-mapping solutions required manual curation. Substantial work would be required to manually curate all atom-mappings, however may be necessary to probe deeper into secondary metabolism. The objectives in the development of this new tool included unbiased assessment of isotopologue sources at genome-scale, however ultimately became subject to biased path-searching in central

energy metabolism. This current implementation has considerable room for improvement, however is presented here as a foundation for additional experiments and simulations.

This algorithm, despite the sheer complexity uncovered, still represents a relatively simplistic purview, as only one labeled reactant is considered per reaction and does not include bimolecular combinations. Additionally, while the LC-MS method utilized in this platform performs fragmentation analysis, this data is essentially discarded given the incomplete positional carbon information. While isotopomer analysis would greatly specify enzyme activities over isotopologues and trim many of the IsoPathFinder possibilities, significant technological advancements are needed to observe complete isotopomer distributions at a systems-level. The current version of this algorithm does store results for each isotopomer separately, before converting labeled SMILES names to common isotopologue names. Consequently, small modifications to the program would allow path-tracing for any isotopomer of interest, and may initially serve as a purely *in silico* approach to simulate metabolic networks with greater specificity. However, positional information from LC-MS or NMR datasets will require exact isotopomer matching to SMILES formats before integration with the IsoPathFinder simulations. As an intermediate step, partial information on positional labeling can be currently incorporated into the IsoPathFinder program. For instance, the alanine M+2 isotopologue has three possible isotopomers. Under the current LC-MS platform, alanine $M+2_{+2}$ represents one of these isotopomers, while alanine $M+2_{+1}$ represents the remaining two isotopomers. The IsoPathFinder can be adjusted to consider any subset of isotopomer distributions.

The greatest potential for metabolic discovery with this approach likely resides in synergy between expanded experimental designs, metabolomic analyses, and IsoPathFinder-derived paths from non-glucose sources. As an example, the data from Chapter 4 prompted simulations with a $^{13}C_3$ serine input tracer to define a set of unique isotopologues which can guide additional experimentation to discover the fate of the cyclic carbon flux into serine M+3 (Figure 5.8). Of potentially greater benefit to metabolic discovery, *in silico* simulations of labeling possibilities can

rationalize novel isotope tracer experiments *a priori* and guide metabolomic methods towards the biological questions of interest.

Table 5.1 IsoPathFinder paths nine reactions deep from $^{13}C_6$ glucose to serine M+3 after trimming procedures

Reaction	1	2
	alpha-D-Glucose M+6	alpha-D-Glucose M+6
1	['EC-4.3.1.17']	['EC-4.3.1.17']
	Hexose-6-Phosphate M+6	Hexose-1-Phosphate M+6
2	['EC-2.7.1.1', 'EC-2.7.1.147']	['EC-3.1.3.10','EC-2.7.1.6','EC-2.4.1.231']
	1,6-di-o-phosphonohexopyranose M+6	1,6-di-o-phosphonohexopyranose M+6
3	['EC-2.7.1.11']	['EC-2.7.1.10']
	Glycerophosphoric acid M+3	Glycerophosphoric acid M+3
4	['EC-4.1.2.13']	['EC-4.1.2.13']
	1,3-Bisphosphoglycerate M+3	1,3-Bisphosphoglycerate M+3
5	['EC-1.2.1.12']	['EC-1.2.1.12']
	3-Phosphoglycerate M+3	3-Phosphoglycerate M+3
6	['EC-2.7.2.3']	['EC-2.7.2.3']
	2-phosphoglyceric acid M+3	2-phosphoglyceric acid M+3
7	['EC-5.4.2.12']	['EC-5.4.2.12']
	phosphoenolpyruvate M+3	phosphoenolpyruvate M+3
8	['EC-4.2.1.11']	['EC-4.2.1.11']
	Pyruvate M+3	Pyruvate M+3
9	['EC-2.7.1.40']	['EC-2.7.1.40']
	Serine M+3	Serine M+3
Gibbs Sum (kJ)	-47.73306331	-30.18725196
Number of Circadian ECs	1	1
Circadian ECs	['EC-2.7.1.40']	['EC-2.7.1.40']

Table 5.2 IsoPathFinder paths fourteen reactions deep from $^{13}C_6$ glucose to glutamine M+2 after trimming procedures. Note not all ECs are displayed for the sake of brevity

Reaction	1	2
	alpha-D-Glucose M+6	alpha-D-Glucose M+6
1	['EC-2.7.1.1', 'EC-2.7.1.147']	['EC-2.7.1.1', 'EC-2.7.1.147']
	Hexose-6-Phosphate M+6	Hexose-6-Phosphate M+6
2	['EC-2.7.1.11']	['EC-2.7.1.11']
	1,6-di-o-phosphonohexopyranose M+6	1,6-di-o-phosphonohexopyranose M+6
3	['EC-4.1.2.13']	['EC-4.1.2.13']
	Glycerophosphoric acid M+3	Glycerophosphoric acid M+3
4	['EC-1.2.1.12']	['EC-1.2.1.12']
	1,3-Bisphosphoglycerate M+3	1,3-Bisphosphoglycerate M+3
5	['EC-2.7.2.3']	['EC-2.7.2.3']
	3-Phosphoglycerate M+3	3-Phosphoglycerate M+3
6	['EC-5.4.2.12']	['EC-5.4.2.12']
	2-phosphoglyceric acid M+3	2-phosphoglyceric acid M+3
7	['EC-4.2.1.11']	['EC-4.2.1.11']
	phosphoenolpyruvate M+3	phosphoenolpyruvate M+3
8	['EC-2.7.1.40']	['EC-2.7.1.40']
	Pyruvate M+3	Pyruvate M+3
9	['EC-1.2.4.1', 'EC-1.2.1', 'EC-1.2.7.1', 'EC-2.3.1.54']	['EC-2.2.1.6']
	Acetyl-CoA M+2	Acetohydroxybutyrate M+2
10	['EC-2.3.3.8','EC-2.3.3.1']	['EC-2.2.1.6']
	Citrate M+2	Pyruvate M+2
11	['EC-4.2.1.3']	['EC-4.1.3.16']
	Isocitrate M+2	2-Hydroxy-4-oxopentanedioic acid M+2
12	['EC-1.1.1.42', 'EC-1.1.1.41']	['EC-2.6.1.23']
	alpha-ketoglutarate M+2	alpha-ketoglutarate M+2
13	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]
	Glutamate M+2	Glutamate M+2
14	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]
	Glutamine M+2	Glutamine M+2
Gibbs Sum (kJ)	-297.2240567	-293.8480185
Number of Circadian ECs	6	3
Circadian ECs	['EC-4.2.1.3', 'EC-2.7.1.40', 'EC-2.3.3.1', 'EC-2.6.1.13',...]	['EC-2.6.1.16', 'EC-2.7.1.40', 'EC-2.6.1.13']

Reaction	3	4
	alpha-D-Glucose M+6	alpha-D-Glucose M+6
1	['EC-2.7.1.1', 'EC-2.7.1.147']	['EC-2.7.1.1', 'EC-2.7.1.147']
	Hexose-6-Phosphate M+6	Hexose-6-Phosphate M+6
2	['EC-2.7.1.11']	['EC-2.7.1.11']
	1,6-di-o-phosphonohexopyranose M+6	1,6-di-o-phosphonohexopyranose M+6
3	['EC-4.1.2.13']	['EC-4.1.2.13']
	Glycerophosphoric acid M+3	Glycerophosphoric acid M+3
4	['EC-1.2.1.12']	['EC-1.2.1.12']
	1,3-Bisphosphoglycerate M+3	1,3-Bisphosphoglycerate M+3
5	['EC-2.7.2.3']	['EC-2.7.2.3']
	3-Phosphoglycerate M+3	3-Phosphoglycerate M+3
6	['EC-5.4.2.12']	['EC-5.4.2.12']
	2-phosphoglyceric acid M+3	2-phosphoglyceric acid M+3
7	['EC-4.2.1.11']	['EC-4.2.1.11']
	phosphoenolpyruvate M+3	phosphoenolpyruvate M+3
8	['EC-2.7.1.40']	['EC-2.7.1.40']
	Pyruvate M+3	Pyruvate M+3
9	['EC-1.2.4.1', 'EC-1.2.1', 'EC-1.2.7.1', 'EC-2.3.1.54']	['EC-6.4.1.1', 'EC-2.1.3.1', 'EC-1.1.1.38']
	Acetyl-CoA M+2	Oxaloacetate M+3
10	['EC-1.2.1', 'EC-1.2.7.1', 'EC-2.3.1.54', 'EC-1.2.4.1']	['EC-2.3.3.1']
	Pyruvate M+2	Citrate M+3
11	['EC-4.1.3.16']	['EC-4.2.1.3']
	2-Hydroxy-4-oxopentanedioic acid M+2	Isocitrate M+3
12	['EC-2.6.1.23']	['EC-1.1.1.42', 'EC-1.1.1.41']
	alpha-ketoglutarate M+2	alpha-ketoglutarate M+2
13	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]
	Glutamate M+2	Glutamate M+2
14	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]
	Glutamine M+2	Glutamine M+2
Gibbs Sum (kJ)	-293.8480185	-290.5400044
Number of Circadian ECs	3	7
Circadian ECs	['EC-2.6.1.16', 'EC-2.7.1.40', 'EC-2.6.1.13']	['EC-4.2.1.3', 'EC-2.3.3.1', 'EC-2.7.1.40', 'EC-6.4.1.1',...]

Reaction	5	6
	alpha-D-Glucose M+6	alpha-D-Glucose M+6
1	['EC-3.1.3.10','EC-2.7.1.6','EC-2.4.1.231']	['EC-3.1.3.10','EC-2.7.1.6','EC-2.4.1.231']
	Hexose-1-Phosphate M+6	Hexose-1-Phosphate M+6
2	['EC-2.7.1.10']	['EC-2.7.1.10']
	1,6-di-o-phosphonohexopyranose M+6	1,6-di-o-phosphonohexopyranose M+6
3	['EC-4.1.2.13']	['EC-4.1.2.13']
	Glycerophosphoric acid M+3	Glycerophosphoric acid M+3
4	['EC-1.2.1.12']	['EC-1.2.1.12']
	1,3-Bisphosphoglycerate M+3	1,3-Bisphosphoglycerate M+3
5	['EC-2.7.2.3']	['EC-2.7.2.3']
	3-Phosphoglycerate M+3	3-Phosphoglycerate M+3
6	['EC-5.4.2.12']	['EC-5.4.2.12']
	2-phosphoglyceric acid M+3	2-phosphoglyceric acid M+3
7	['EC-4.2.1.11']	['EC-4.2.1.11']
	phosphoenolpyruvate M+3	phosphoenolpyruvate M+3
8	['EC-2.7.1.40']	['EC-2.7.1.40']
	Pyruvate M+3	Pyruvate M+3
9	['EC-1.2.4.1', 'EC-1.2.1', 'EC-1.2.7.1', 'EC-2.3.1.54']	['EC-1.2.4.1', 'EC-1.2.1', 'EC-1.2.7.1', 'EC-2.3.1.54']
	Acetyl-CoA M+2	Acetyl-CoA M+2
10	['EC-2.3.3.1','EC-2.3.3.8']	['EC-1.2.1', 'EC-1.2.7.1', 'EC-2.3.1.54', 'EC-1.2.4.1']
	Citrate M+2	Pyruvate M+2
11	['EC-4.2.1.3']	['EC-4.1.3.16']
	Isocitrate M+2	2-Hydroxy-4-oxopentanedioic acid M+2
12	['EC-1.1.1.42', 'EC-1.1.1.41']	['EC-2.6.1.23']
	alpha-ketoglutarate M+2	alpha-ketoglutarate M+2
13	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]
	Glutamate M+2	Glutamate M+2
14	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]
	Glutamine M+2	Glutamine M+2
Gibbs Sum (kJ)	-279.6782453	-276.3022071
Number of Circadian ECs	6	3
Circadian ECs	['EC-4.2.1.3', 'EC-2.7.1.40', 'EC-2.3.3.1', 'EC-2.6.1.13',...]	['EC-4.2.1.3', 'EC-2.3.3.1', 'EC-2.7.1.40', 'EC-6.4.1.1',...]

Reaction	7	8
	alpha-D-Glucose M+6	alpha-D-Glucose M+6
1	['EC-3.1.3.10','EC-2.7.1.6','EC-2.4.1.231']	['EC-3.1.3.10','EC-2.7.1.6','EC-2.4.1.231']
	Hexose-1-Phosphate M+6	Hexose-1-Phosphate M+6
2	['EC-2.7.1.10']	['EC-2.7.1.10']
	1,6-di-o-phosphonohexopyranose M+6	1,6-di-o-phosphonohexopyranose M+6
3	['EC-4.1.2.13']	['EC-4.1.2.13']
	Glycerophosphoric acid M+3	Glycerophosphoric acid M+3
4	['EC-1.2.1.12']	['EC-1.2.1.12']
	1,3-Bisphosphoglycerate M+3	1,3-Bisphosphoglycerate M+3
5	['EC-2.7.2.3']	['EC-2.7.2.3']
	3-Phosphoglycerate M+3	3-Phosphoglycerate M+3
6	['EC-5.4.2.12']	['EC-5.4.2.12']
	2-phosphoglyceric acid M+3	2-phosphoglyceric acid M+3
7	['EC-4.2.1.11']	['EC-4.2.1.11']
	phosphoenolpyruvate M+3	phosphoenolpyruvate M+3
8	['EC-2.7.1.40']	['EC-2.7.1.40']
	Pyruvate M+3	Pyruvate M+3
9	['EC-2.2.1.6']	['EC-6.4.1.1', 'EC-2.1.3.1', 'EC-1.1.1.38']
	Acetohydroxybutyrate M+2	Oxaloacetate M+3
10	['EC-2.2.1.6']	['EC-2.3.3.1']
	Pyruvate M+2	Citrate M+3
11	['EC-4.1.3.16']	['EC-4.2.1.3']
	2-Hydroxy-4-oxopentanedioic acid M+2	Isocitrate M+3
12	['EC-2.6.1.23']	['EC-1.1.1.42', 'EC-1.1.1.41']
	alpha-ketoglutarate M+2	alpha-ketoglutarate M+2
13	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]	['EC-2.6.1.39', 'EC-1.4.1.3', 'EC-1.4.1.13', 'EC-2.6.1',...]
	Glutamate M+2	Glutamate M+2
14	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]	['EC-2.6.1.85', 'EC-6.3.5.4', 'EC-2.4.2.14', 'EC-1.4.1.13',...]
	Glutamine M+2	Glutamine M+2
Gibbs Sum (kJ)	-276.3022071	-272.9941931
Number of Circadian ECs	3	7
Circadian ECs	['EC-2.6.1.16', 'EC-2.7.1.40', 'EC-2.6.1.13']	['EC-4.2.1.3', 'EC-2.3.3.1', 'EC-2.7.1.40', 'EC-6.4.1.1',...]

Table 5.3 IsoPathFinder paths eleven reactions deep from $^{13}C_6$ glucose to reduced glutathione M+2 after trimming procedures

Reaction	1	2
	alpha-D-Glucose M+6	alpha-D-Glucose M+6
1	['EC-2.7.1.1', 'EC-2.7.1.147']	['EC-3.1.3.10','EC-2.7.1.6','EC-2.4.1.231']
	Hexose-6-Phosphate M+6	Hexose-1-Phosphate M+6
2	['EC-2.7.1.11']	['EC-2.7.1.10']
	1,6-di-o-phosphonohexopyranose M+6	1,6-di-o-phosphonohexopyranose M+6
3	['EC-4.1.2.13']	['EC-4.1.2.13']
	Glycerophosphoric acid M+3	Glycerophosphoric acid M+3
4	['EC-1.2.1.12']	['EC-1.2.1.12']
	1,3-Bisphosphoglycerate M+3	1,3-Bisphosphoglycerate M+3
5	['EC-2.7.2.3']	['EC-2.7.2.3']
	3-Phosphoglycerate M+3	3-Phosphoglycerate M+3
6	['EC-5.4.2.12']	['EC-5.4.2.12']
	2-phosphoglyceric acid M+3	2-phosphoglyceric acid M+3
7	['EC-4.2.1.11']	['EC-4.2.1.11']
	phosphoenolpyruvate M+3	phosphoenolpyruvate M+3
8	['EC-2.7.1.40']	['EC-2.7.1.40']
	Pyruvate M+3	Pyruvate M+3
9	['EC-4.3.1.17']	['EC-4.3.1.17']
	Serine M+3	Serine M+3
10	['EC-2.1.2.1', 'EC-2.1.2']	['EC-2.1.2.1', 'EC-2.1.2']
	Glycine M+2	Glycine M+2
11	['EC-6.3.2.3']	['EC-6.3.2.3']
	Glutathione Reduced M+2	Glutathione Reduced M+2
Gibbs Sum (kJ)	-82.5341937	-64.98838235
Number of Circadian ECs	1	1
Circadian ECs	['EC-2.7.1.40']	['EC-2.7.1.40']

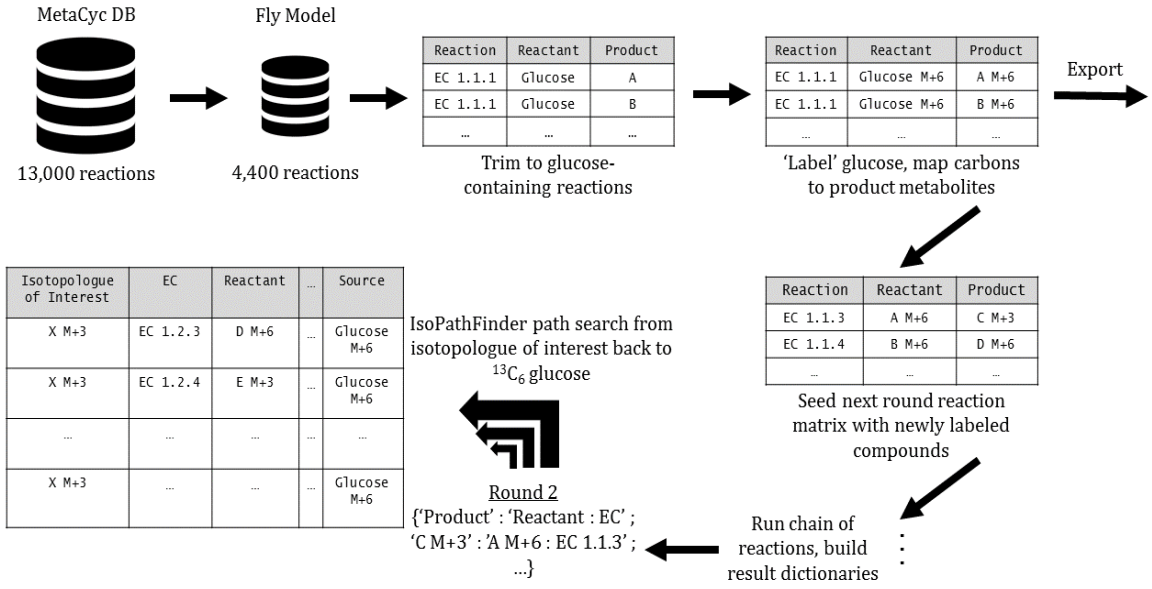


Figure 5.1 Workflow for genome-scale reaction modeling to trace isotopologue paths from $^{13}C_6$ glucose. Atom-mapping solutions for *Drosophila* reactions in the MetaCyc database are trimmed to reactions which contain glucose. Those glucose carbons are then 'labeled' and mapped to carbons in the product metabolites. These results are exported, followed by an additional reaction round using the newly labeled metabolites, until all reactions are calculated. These results can be strung together to build paths from the isotopologue of interest back to the $^{13}C_6$ glucose tracer in Round 1 in the IsoPathFinder algorithm.

A

```
In [48]: %%time

         #Round 1

         #Find all reactions containing list of labeled compounds
         NextRoundReaction=BuildReactionMatrixFromCpdList(LabListInitial)

         #Run the reaction
         NextRoundReaction.Compounds=NextRoundReaction.Compounds.apply(GenericLabelingReactionMatrixApply)
         NextRoundReaction=UnpackAndFill(NextRoundReaction)
         NextRoundReaction['NewCpds']=NextRoundReaction.Products

         #Get list of newly labeled products for the next round
         NextRoundCpdList=NewLabeledCpdList(NextRoundReaction)

         #Write to file if list of labeled compounds is to be used later to restart the simulations
         CpdFile='LabeledCpds_FromRound_1.csv'
         with open(CpdFile,'w') as output:
             writer=csv.writer(output,lineterminator='\n')
             for val in NextRoundCpdList:
                 writer.writerow([val])

         #Build the results in a dictionary format and write to file
         DictionaryRound=MakingDict(NextRoundReaction)
         output=open('Dictionary_FromRound_1.pkl','wb', -1)
         pickle.dump(DictionaryRound,output,protocol=4)

         print('Round 1 Done')

Round 1 Done
CPU times: user 64 ms, sys: 0 ns, total: 64 ms
Wall time: 63.8 ms
```

B

```
In [42]: DictionaryRound

Out[42]: defaultdict(<function __main__.MakingDict.<locals>.<lambda>>,
                     {'(3,4,5-trihydroxy-6-oxotetrahydro-2h-pyran-2-yl)methyldihydrogen-phosphat M+6': defaultdict(list,
                                 {'6-o-phosphonohexopyranose M+6': ['EC-1.1.1.363',
                                   'EC-1.1.98.2']}),
                      '1,6-di-o-phosphonohexopyranose M+6': defaultdict(list,
                                 {'1-O-phosphonohexopyranose M+6': ['EC-2.7.1.106',
                                   'EC-2.7.1.10',
                                   'RXN-16997',
                                   'RXN-16995'],
                                  '6-o-phosphonohexopyranose M+6': ['RXN-16996',
                                   'RXN-16998']}),
                      '1-O-phosphonohexopyranose M+6': defaultdict(list,
                                 {'1-O-phosphonohexopyranose M+6': ['RXN0-5108',
                                   'RXN-10770',
                                   'RXN-10639',
                                   'EC-2.4.1.64'],
                                  '6-o-phosphonohexopyranose M+6': ['EC-5.4.2.5',
                                   'EC-5.4.2.6'],
                                  'D-(+)-Trehalose M+6': ['EC-2.4.1.64',
                                   'EC-2.4.1.231'],
```

Product

Reactants

Enzymes

C

```
sethpy@sethpy-VirtualBox:~/Documents/CarbonTracing/FlyMetaCycResults$ ipython
Python 3.5.2 |Anaconda 4.2.0 (64-bit)| (default, Jul  2 2016, 17:53:06)
Type "copyright", "credits" or "license" for more information.

IPython 5.1.0 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: %run IsoPathFinder.ipynb
Welcome to IsoPathFinder Version 1.0, to calculate isotopomer paths from 13C6 Gl
ucose, type BuildPaths(), followed by the isotopmer of interest

In [2]: BuildPaths()
Enter isotopomer, or type "help" for lists of common isotopomers (then call Buil
dPaths again with specified isotopomer of interest): Serine M+3
Isotopomer found, calculating results... this may take a few minutes
Successfully run, paths stored as csv in local directory

In [3]:
```

Figure 5.2 (A) Code example to simulate the first set of reactions from $^{13}C_6$ glucose, starting from a subset of reactions which contain glucose, to carbon-labeling of reaction products, followed by exporting results in a Python dictionary and a list of newly labeled compounds to seed the next reaction round. (B) Example dictionary, where the keys consist of product metabolites which received a ^{13}C label, and the values are the reactants and enzymes which catalyzed the reaction. (C) Example of user input to the IsoPathFinder algorithm to build paths to serine M+3, using the dictionaries depicted in (B).

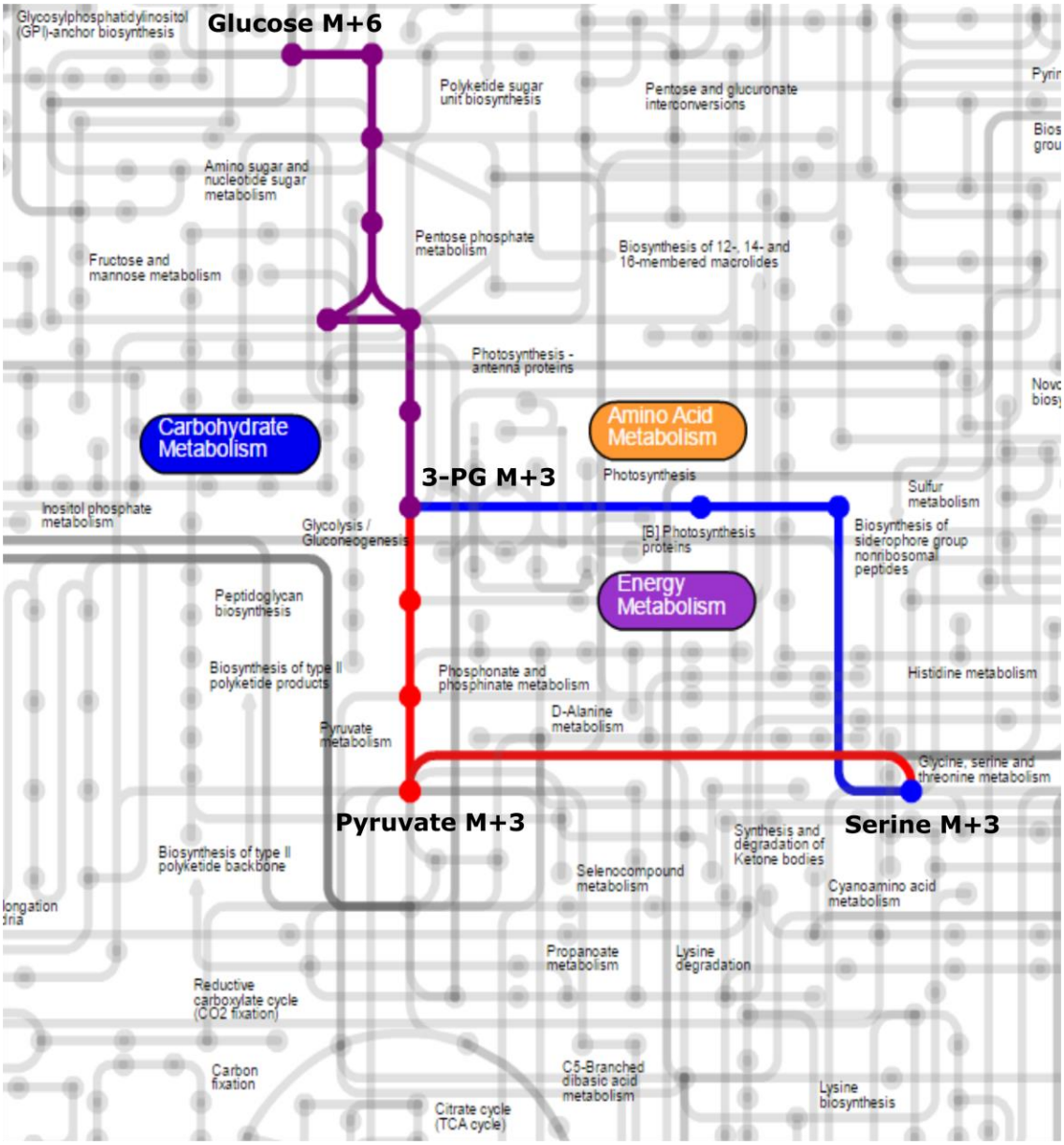


Figure 5.3 Two major paths to serine M+3 from glucose M+6, calculated by the IsoPathFinder. The canonical serine biosynthesis pathway (blue) did not contain any cyclic enzymes and yielded a positive ΔG, while the alternative path (red) through pyruvate M+3 contained a cyclic enzyme based on previous expression datasets, and was more thermodynamically favorable. Diagrams generated using iPath2 (Yamada et al., 2011).

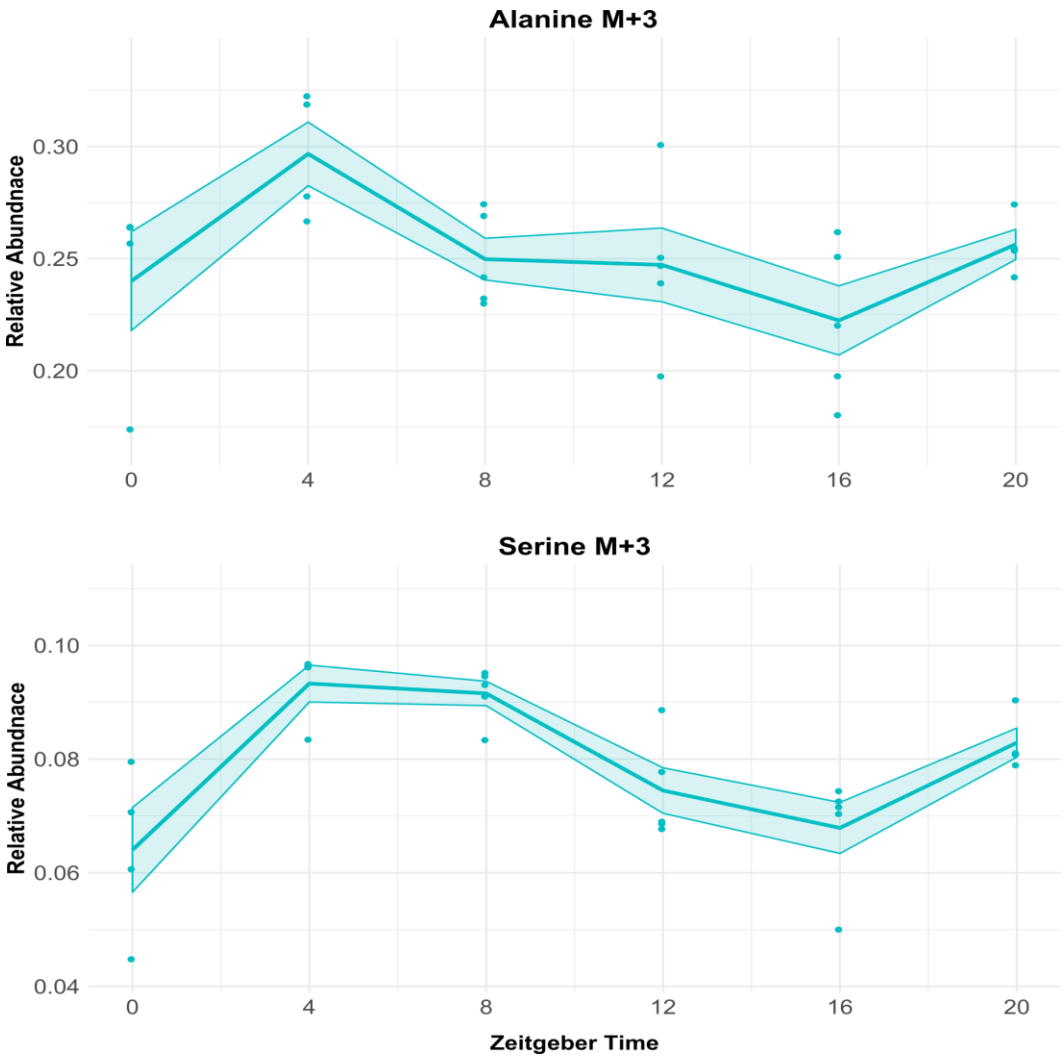


Figure 5.4 Temporal overlap of alanine and serine M+3 relative abundances in WT flies, from isotope tracer experiments described in Chapter 4, which was revealed through exploration of longer paths to serine M+3.

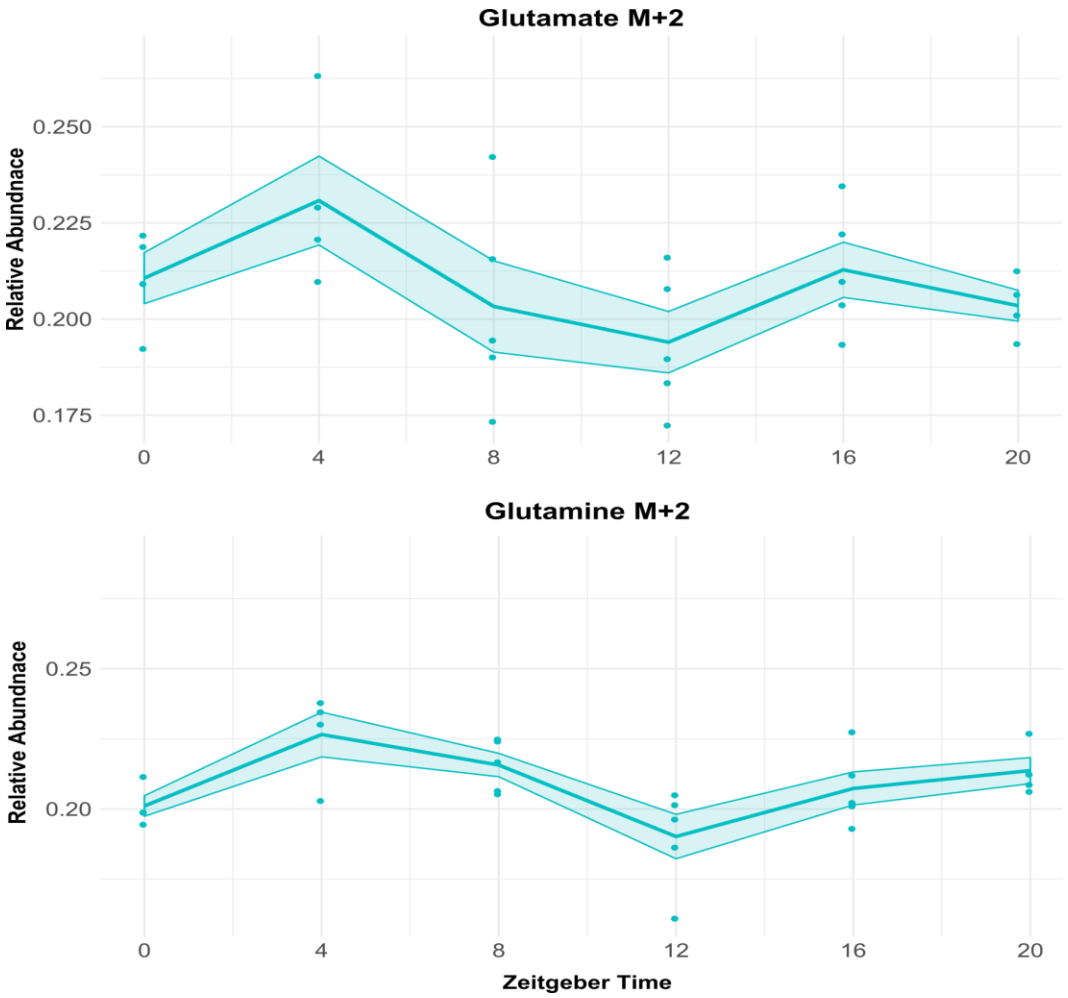


Figure 5.5 Temporal overlap of glutamine and glutamate M+2 relative abundances in WT flies, from isotope tracer experiments described in Chapter 4. All paths to glutamine M+2 from IsoPathFinder analysis included glutamate M+2 in the preceding step.

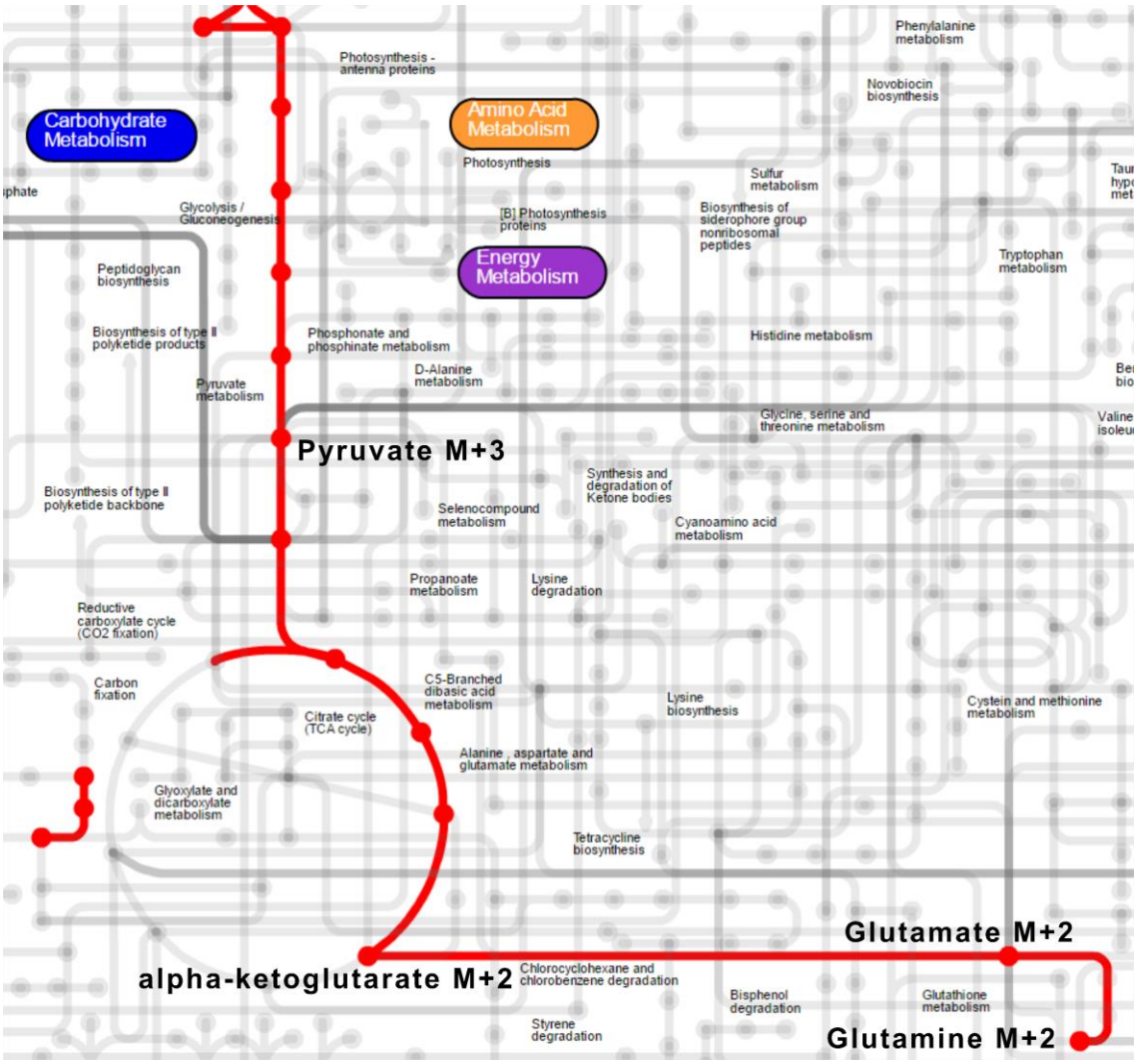


Figure 5.6 Most likely path to glutamine M+2 from glucose M+6 through the TCA cycle, calculated by the IsoPathFinder and trimmed by thermodynamic favorability and circadian enzyme expression from previous studies. Diagrams generated using iPath2 (Yamada et al., 2011).

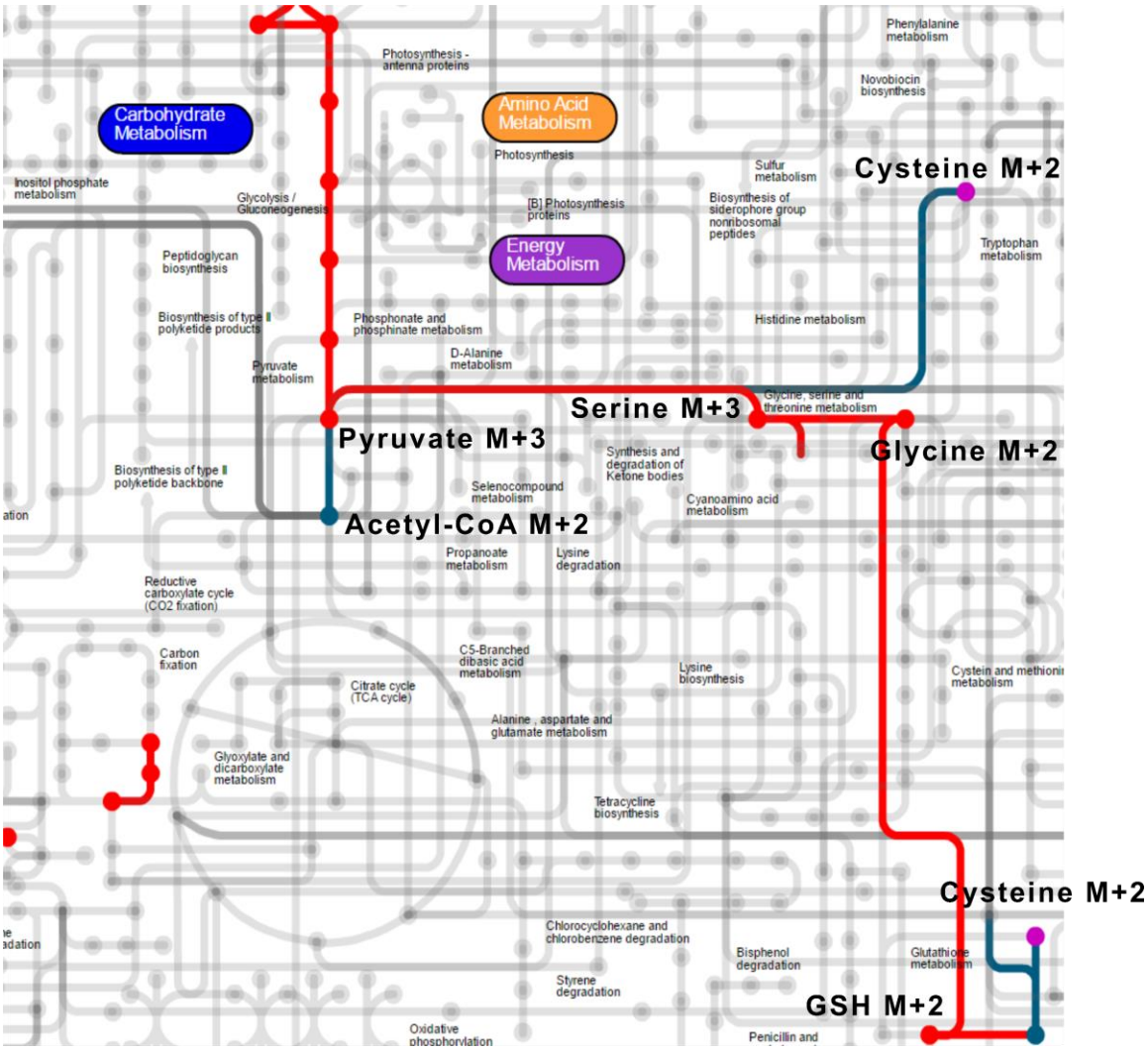


Figure 5.7 Most likely and shortest path to GSH M+2 from glucose M+6 through glycine M+2, calculated by the IsoPathFinder (red). Longer paths were calculated through cysteine M+2 (green). Note cysteine M+2 itself could not be connected to glutathione production with this KEGG-based visualization tool (pink). Diagrams generated using iPath2 (Yamada et al., 2011).

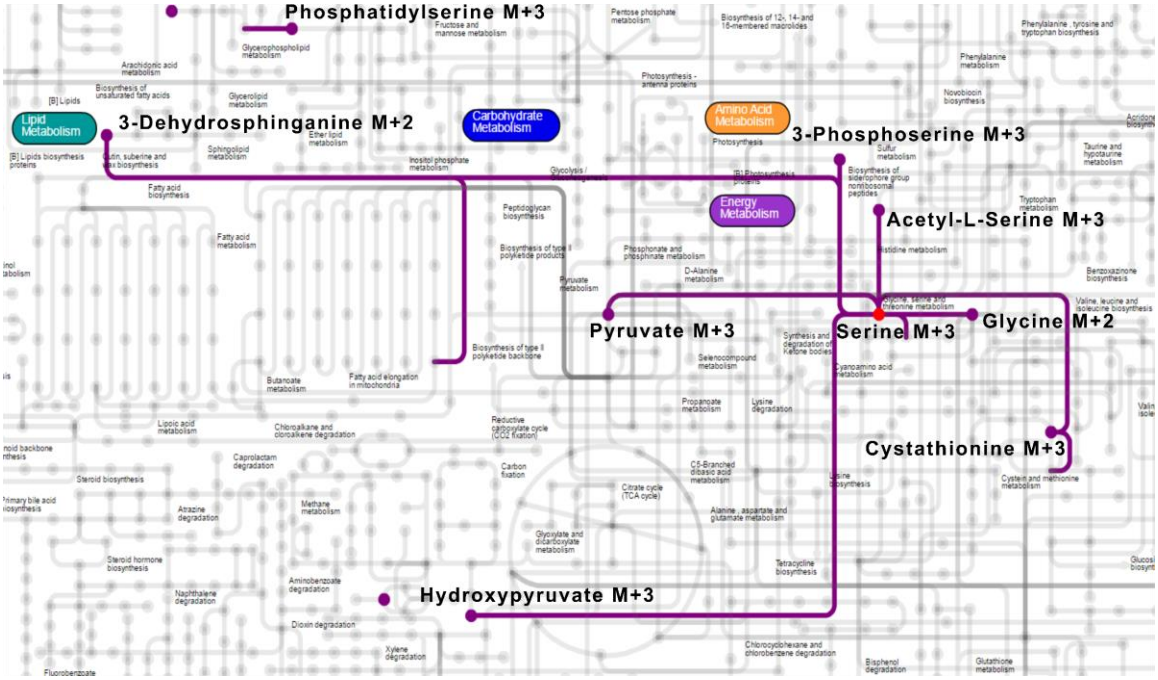


Figure 5.8 Isotopologues one reaction removed from a simulated $^{13}C_3$ serine tracer (red), proposing additional experiments and LC-MS analyses to confirm the fate of cyclic carbon flux into serine from $^{13}C_6$ glucose noted in Chapter 4. Diagrams generated using iPath2 (Yamada et al., 2011).

CHAPTER 6 – Overview

6.1 Project Summary

The motivation for this work stems from our incomplete understanding of the intersection of clocks, sleep, and metabolism, and the realization that new tools are necessary to unearth these connections and track inherently dynamic physiological processes. While gross metabolic parameters such as glucose homeostasis have been extensively studied vis-à-vis disrupted circadian rhythms and sleep, mechanistic detail cannot be obtained from these measures. Considerably deeper and more complex analyses are required to uncover the extent of diurnal variance in metabolism. Metabolomics has recently emerged as an approach to holistically probe metabolic processes and integrate both endogenous and exogenous influences on host physiology. As with other high-throughput 'omic technologies, metabolomic analyses rely on instrumentation with multiple tunable parameters which impacts the accurate detection of numerous variables, and requires efficient and rational optimization procedures. In Chapter 2, new LC-MS metabolomics methodology was developed in an iterative, data-driven manner. Design of experiments is commonly employed in process design and engineering, however has never been leveraged to build new LC-MS metabolomics methodology.

These methods can facilitate holistic metabolite profiling in appropriate models of circadian rhythms, highlighted in Chapter 3 by metabolomic analyses in *Drosophila*. In this project, we discovered a dependency of circadian metabolite rhythms on intact genetic clocks, and a set of metabolite rhythms driven by environmental cues. Given cycles in physiology arise from a complex interaction of environmental cues and genetic factors, this study represents an important novel paradigm which separates the impact of daily light cycles and intrinsic clocks on metabolism. Additionally, metabolomic profiling in *Drosophila* at sufficiently high sampling resolution necessary for ultradian rhythm analysis had not been previously performed, and advances flies as a model to uncover new basic mechanisms in chronometabolism.

LC-MS metabolomics methods can thus be effectively developed to discern metabolic cycles, however these methods do not provide clear insight into reaction kinetics, presenting an additional challenge in obtaining cogent metabolic mechanisms. The LC-MS methods described in Chapter 2 were modified to address this limitation through isotope enrichment detection from a novel $^{13}C_6$ glucose bolus injection platform in *Drosophila*, described in Chapter 4. Current technological limitations prevent quantitative metabolic flux modeling across a circadian time-series *in vivo*, however deep isotopolomic profiling provided relative pathway fluxes in amino acid metabolism and redox across circadian time and sleep status. This approach not only represents a new isotope-tracing tracing paradigm in *Drosophila*, but also an unprecedented scope of isotopologue analysis over prior metabolic flux study designs. Multivariate modeling on isotopologues had not been previously utilized to obtain a fluxomics assessment of an *in vivo* system, which revealed a global shift in carbohydrate oxidation early in the active phase of flies. This platform can have broad applicability in defining the metabolic consequence of clock and sleep disruption, as well as translatable experimental designs regarding nutrient processing and feeding paradigms to promote metabolic health.

The interpretability of systems-level ‘omics analyses lag behind advancements in instrumentation. While Chapter 4 outlines a novel approach to gauge metabolic flux, the source of many of these isotopologues cannot be fully explained by predefined biochemical pathways. The new metabolic variables discovered in Chapter 4 require a new means of interpretation, presenting yet another challenge in deriving true metabolic phenotypes. Fortunately, with the advent of genome-scale reaction annotation, all reactions can be considered which produce these detectable isotopologues. A new computational approach is developed which enumerates all possible isotope transfer events with atomic detail from an input $^{13}C_6$ glucose tracer, and subsequently reconstructs all paths which connects the input tracer to the user-defined isotopologue of interest, outlined in Chapter 5. As a proof of concept, this algorithm calculated paths to notable isotopologues obtained in Chapter 4, including serine M+3, glutamine M+2, and GSH M+2. After a series of rational trimming procedures, a subset of those possible routes

remained, which provide supporting evidence for suspected sources of each of these isotopologues, as well as propose alternative mechanisms that warrant further experimentation. This tool simulated downstream isotopologues from a $^{13}C_3$ serine tracer, as an example of the potential utility to guide the researcher to test specific hypotheses through appropriate tracer experiments and metabolomics methodology. This approach represents a new hypothesis-generating bioinformatic tool built in an open-source platform for widespread use. Notably, the extent of possibilities calculated with this algorithm presents a degree of complexity which may require a shift in perspectives to consider metabolomic datasets in terms of updated probabilities, as attributing observable metabolite data to a single source or reaction is likely to be false. As technology and computational tools continue to improve, a new era of discovery can form to truly shed light on the complex nature of chronometabolism. The efforts outlined in this thesis work formed a logical progression of problem-solving endeavors, ultimately producing new analytical methodology, metabolic modeling algorithms, and most importantly, novel biological insight. Much work remains to improve the largely underdetermined platforms with which we probe metabolism, however this thesis work represents but a small effort to “make measurable what is not so.”

6.2 Future Directions

In this work, there exist a multitude of opportunities to both improve methodologies and provide additional biological insight. The DoE approach outlined in Chapter 2 not only produced new metabolomics methodologies, which drove much of the results obtained in Chapters 3 and 4, but also itself addressed a need to rapidly develop new methods in a data-driven manner. This approach is important, as no single method is adequate to detect the entire metabolome, and certain subsets of metabolites which cannot be detected with our current methods may be necessary to test additional hypotheses. For example, improved detection methods for sugars and would add clarity to the results obtained in Chapter 3, while improved sensitivity of amino acid isotopomers would greatly enhance reaction specificity for the predictions generated in Chapters 4 and 5. Algorithmic improvements to the IsoPathFinder approach, alongside improved

database curation and LC-MS methods, can also augment this new platform of isotope-based metabolic discovery.

The approaches presented in this work also provide ample space for novel chronometabolic insight with translatable potential. Future studies from Chapter 3 should seek to determine the clock-dependent role of *per* in circadian metabolism through analysis of other genetic clock models, and separate the impact of clocks and feeding rhythms on metabolite cycles. Perturbed metabolic homeostasis has been noted in mice with dysfunctional clocks, which may be alleviated through certain behavioral and feeding paradigms. Given the ease of genetic manipulation in flies, and the relatively low cost of experimentation, our ability to improve health outcomes amidst altered sleep and feeding cycles with practical benefits for human health may be augmented by future *Drosophila* metabolomic studies. Simply acquiring metabolomic data from the combinatorial explosion of genetic mutants, environmental influences, and feeding paradigms will likely lead to an uninterpretable excess of data. However, metabolomics may serve as a purview of underlying mechanisms for experiments that aim to improve metabolic outcomes under paradigms of desynchrony. These experiments may also benefit from an altered perspective in the utility of metabolomics and systems-biology research. One major goal in biomedical research is to improve outcomes. If one is to think of metabolomics as a metabolic fingerprint of a biological system, a prevailing objective should be to define the biometric fingerprint, and derive means to shift a diseased fingerprint towards a healthy one. With this perspective in mind, metabolomics may serve to gauge the success of efforts to restore metabolic homeostasis under conditions of altered feeding and activity patterns.

Apropos to feeding, the isotope tracer approach in Chapter 4 can more directly answer questions regarding diurnal variation in nutrient oxidation and reduction, and may provide detailed mechanisms to the consequence of dysfunctional clocks and the benefits of time-restricted feeding. The platform itself will benefit from additional metabolomics methods, as only a relatively small number of metabolites are analyzed in the current analysis given the number of isotopomer MRMs required per metabolite. Experiments which seek to test specific hypotheses, for example

cycles in isotopologue products of serine M+3, will need to modify the existing method to accurately detect all necessary isotopologues with greater sensitivity. Nonetheless, this platform lays a foundation for greatly expanded experimentation, most notably additional isotope tracer injections to probe pathways in amino acids and lipids relevant to circadian metabolism. Additionally, greater reaction specificity will be obtained with proper utilization of fragmentation data, although will require significant methodological improvements. The IsoPathFinder approach can readily assimilate isotopomer data, however would still benefit from algorithmic improvement to search for longer path lengths. Nonetheless, this algorithm can effectively simulate isotope tracer experiments to guide experimental designs in previously unexplored aspects of metabolism, generating a series of future directions too numerous to fully mention here.

6.3 Deeper Metabolic Phenotyping

Even with the relatively simplistic approach in the IsoPathFinder algorithm, the paths enumerated for serine, glutamine and GSH may not represent the full extent of possibilities, as longer paths are easily envisioned. Even after paring unlikely paths through the assimilation of previously published datasets and thermodynamic considerations, the number of possible metabolic routes remains humbling. While scientific efforts typically strive for absolutism in data interpretation, all too often alternative hypotheses are discarded or unconsidered. As we continue to uncover the extent of possibilities in genome-wide networks, our current analytical tools confine us to increasingly underdetermined observable systems. If the IsoPathFinder results remotely represent the extent of possibilities in metabolic systems, we are left with a considerable weight of uncertainty in our experimental results. Advancements in metabolomics technologies may yet meet this challenge, however there are benefits to shifting our philosophy from absolutism to probability distributions. As an example, if the 7246 initial paths to glutamine M+2 are viewed as equally likely, additional data can increase posterior probabilities towards a smaller subset of paths. Even so, attributing experimentally observed results to any one of these paths with high certainty is difficult and perhaps unrealistic. With this perspective in mind, I have sought to

maintain a relatively conservative interpretation of the results obtained in Chapters 4 and 5, as substantial follow-up experimentation would be required to definitively claim the contribution of a particular metabolic pathway to observable organismal phenotypes. Encouragingly, these approaches can be leveraged to test new hypotheses with greater specificity, with the hope that widespread usage of this tool can empower isotope-based metabolic studies in a new era of systems-level biomedical discovery.

In these objectives, I have attempted to shed some light on some unmet needs, both in our understanding of human health, and in deficiencies to define and interpret organismal phenotypes. We currently live in an environment extremely disparate from our origins. Great minds in recent history have driven incredible innovations to enhance the quality of life in developed countries, across medicine, transportation, communication, and information technology. These advancements have ushered in an era of new societal norms, but have quietly come at some cost to our health. Thomas Edison likely would have never imagined the ubiquity of lights at all hours of the night, including those from our handheld devices. The advancements in communication and information technology have granted our work access to personal time early in the morning or late at night, and we have reached a point where sleep restriction from work demands is commonplace, fueling a notion that busy lives are imperative for productivity. These behavioral aberrations correlate closely with disrupted sleep and eating patterns, and consequently desensitize our perceptions of shift work, once considered unethical. Scientists are struggling to clearly define the consequences of these lifestyle shifts on pace with the advancements in technology. Even if only infinitesimally, I hope to add some understanding to the way in which nutrients are metabolized across circadian time, and the impact of restricted sleep on nutrient processing. In the field of biomedical research, tremendous developments have been made across analytical chemistry, mathematics, and computer science to better define ourselves, both in health and disease. I believe we are just scratching the surface in uncovering the genotype-phenotype relationship. Our definitions of human phenotypes must shift away from reductionism and towards systems-level fingerprinting, through a combination of ‘omics, lifestyle,

and real-world evidence datasets. In doing so, we need to maintain an open mind and separate ourselves from the concepts developed in the 20th century, particularly in defining the 'metabolic phenotype'. In my work, I have attempted to remain unbiased in forging new metabolic pathway dynamics from systems-level datasets, and hope the analytical and computational tools presented in this work can be used broadly, or spark the creation of better tools. We must expand the scope of our search beyond what was once thought to be the solved intracellular metabolome, for in metabolism, as in life, '*panta rhei*'.

BIBLIOGRAPHY

Abbondante S, Eckel-Mahan KL, Ceglia NJ, Baldi P, Sassone-Corsi P. Comparative circadian metabolomics reveal differential effects of nutritional challenge in the serum and liver. J Biol Chem. 2016;291(6):2812–28.

Ahmed N, Weidemann MJ. Purine metabolism in promyelocytic HL60 and dimethylsulphoxide-differentiated HL60 cells. Leuk Res. 1994;18(6):441–51.

A J, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, et al. Extraction and GC/MS analysis of the human blood plasma metabolome. Anal Chem. 2005;77(24):8086-94.

Altman BJ, Hsieh AL, Sengupta A, Krishnanaiah SY, Stine ZE, Walton ZE, et al. MYC Disrupts the circadian clock and metabolism in cancer cells. Cell Metab. 2015;22(6):1009–19.

An PN, Yamaguchi M, Bamba T, Fukusaki E. Metabolome analysis of Drosophila melanogaster during embryogenesis. PLoS One. 2014;9(8):e99519.

Ando K, Kripke D, Ancoli-Israel S. Delayed and advanced sleep phase symptoms. Isr J Psychiatry Relat Sci. 2002;39(1):11–8.

Andretic R, van Swinderen B, Greenspan RJ. Dopaminergic modulation of arousal in Drosophila. Curr Biol. 2005;15(13):1165–75.

Ang JE, Revell V, Mann A, Mäntele S, Otway DT, Johnston JD, et al. Identification of human plasma metabolites exhibiting time-of-day variation using an untargeted liquid chromatography-mass spectrometry metabolomic approach. Chronobiol Int. 2012;29(7):868–81.

Archer SN, Robilliard DL, Skene DJ, Smits M, Williams A, Arendt J, et al. A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. Sleep. 2003;26(4):413–5.

Aschoff J. Circadian rhythms in nan. Science. 1965;148(3676):1427–32.

Asher G, Gatfield D, Stratmann M, Reinke H, Dibner C, Kreppel F, et al. SIRT1 Regulates Circadian Clock Gene Expression through PER2 Deacetylation. Cell. 2008;134(2):317–28.

Balbo M, Leproult R, Van Cauter E. Impact of sleep and its disturbances on hypothalamo-pituitary-adrenal axis activity. Int J Endocrinol. 2010;2010:759234.

Barber AF, Erion R, Holmes TC, Sehgal A. Circadian and feeding cues integrate to drive rhythms of physiology in Drosophila insulin-producing cells. Genes Dev. 2016;30(23):2596-606.

Barbieri M, Bonafe M, Franceschi C, Paolisso G. Insulin/IGF-I-signaling pathway: An evolutionarily conserved mechanism of longevity from yeast to humans. Am J Physiol Endocrinol Metab. 2003;285(5):E1064-71.

Bass J, Takahashi J. Circadian integration of metabolism and energetics. Science. 2010;330(6009):1349–54.

Bass J. Circadian topology of metabolism. Nature. 2012;491(7424):348–56.

Basu SS, Mesaros C, Gelhaus SL, Blair IA. Stable isotope labeling by essential nutrients in cell culture for preparation of labeled coenzyme A and its thioesters. Anal Chem. 2011;83(4):1363-9.

Beaver LM, Klichko VI, Chow ES, Kotwica-Rolinska J, Williamson M, Orr WC, et al. Circadian regulation of glutathione levels and biosynthesis in Drosophila melanogaster. PLoS One. 2012;7(11):e50454.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57(1):289–300.

Bharucha KN. The epicurean fly: Using Drosophila melanogaster to study metabolism. Pediatr Res. 2009;65(2):132–7.

Birse RT, Choi J, Reardon K, Rodriguez J, Graham S, Diop S, et al. High-fat-diet-induced obesity and heart dysfunction are regulated by the TOR pathway in Drosophila. Cell Metab. 2010;12(5):533–44.

Blanksby S, Mitchell T. Advances in mass spectrometry for lipidomics. Annu Rev Anal Chem. 2010;3:433–65.

Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37(8):911-7.

Bolli GB, de Feo P, de Cosmo S, Perriello G, Ventura MM, Calcinaro F, et al. Demonstration of a dawn phenomenon in normal human volunteers. Diabetes. 1984;33(12):1150–3.

Bonneau R. Learning biological networks: From modules to dynamics. Nat Chem Biol. 2008;4(11):658–64.

Borbely AA. A two process model of sleep regulation. Hum Neurobiol. 1982;1(3):195–204.

Bordbar A, McCloskey D, Zielinski DC, Sonnenschein N, Jamshidi N, Palsson BØ. Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. Cell Syst. 2015;1(4):283–92.

Bordbar A, Monk JM, King ZA, Palsson BØ. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet. 2014;15(2):107–20.

Brown, SA. Circadian metabolism: From mechanisms to metabolomics and medicine. Trends Endocrinol Metab. 2016;27(6):415-26.

Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kockhar S, Guy PA. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Anal Chem. 2009;81(9):3285-96.

Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting 13C metabolite labeling patterns from cells. Curr Opin Biotechnol. 2015;34:189–201.

Cajka T, Fiehn O. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. Anal Chem. 2016;88(1):524–45.

Cantó C, Gerhart-Hines Z, Feige JN, Lagouge M, Noriega L, Milne JC, et al. AMPK regulates energy expenditure by modulating NAD+ metabolism and SIRT1 activity. Nature. 2009;458(7241):1056–60.

Capellades J, Navarro M, Samino S, Garcia-Ramirez M, Hernandez C, Simo R, et al. GeoRge: A computational tool to detect the presence of stable isotopelLabeling in LC/MS-based untargeted metabolomics. Anal Chem. 2016;88(1):621–8.

Carpen JD, Von Schantz M, Smits M, Skene DJ, Archer SN. A silent polymorphism in the PER1 gene associates with extreme diurnal preference in humans. J Hum Genet. 2006;51(12):1122–5.

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2016;44(D1):D471–80.

Castro C, Briggs W, Paschos GK, FitzGerald GA, Griffin JL. A metabolomic study of adipose tissue in mice with a disruption of the circadian system. Mol Biosyst. 2015;11(7):1897-906.

Castro-Perez J, Previs SF, McLaren DG, Shah V, Herath K, Bhat G, et al. In vivo D2O labeling to quantify static and dynamic changes in cholesterol and cholesterol esters by high resolution LC/MS. J Lipid Res. 2011;52(1):159-69.

Ceriani MF, Hogenesch JB, Yanovsky M, Panda S, Straume M, Kay SA. Genome-wide expression analysis in Drosophila reveals genes controlling circadian behavior. J Neurosci. 2002;22(21):9305–19.

Chain EB. Landmarks and perspectives in biochemical research. Br Med J. 1965;1(5429):209–20.

Chaix A, Zarrinpar A, Miu P, Panda S. Time-restricted feeding is a preventative and therapeutic intervention against diverse nutritional challenges. Cell Metab. 2014;20(6):991–1005.

Chaneton B, Hillmann P, Zheng L, Martin AC, Maddocks OD, Chokkathukalam A, et al. Serine is a natural ligand and allosteric activator of pyruvate kinase M2. Nature. 2012;491(7424):458-62.

Chintapalli VR, Al Bratty M, Korzekwa D, Watson DG, Dow JT. Mapping an atlas of tissue-specific Drosophila melanogaster metabolomes by high resolution mass spectrometry. PLoS One. 2013;8(10):e78066.

Choi J, Grossbach MT, Antoniewicz MR. Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. Anal Chem. 2012;84(10):4628–32.

Chokkathukalam A, Kim D-H, Barrett MP, Breitling R, Creek DJ. Stable isotope-labeling studies in metabolomics: New insights into structure and dynamics of metabolic networks. Bioanalysis. 2014;6(4):511–4.

Chua EC, Shui G, Lee IT, Lau P, Tan L, Yeo S. Extensive diversity in circadian regulation of plasma lipids and evidence for different circadian metabolic phenotypes in humans. Proc Natl Acad Sci U S A. 2013;110(35):14468–73.

Claridge-Chang A, Wijnen H, Naef F, Boothroyd C, Rajewsky N, Young MW. Circadian regulation of gene expression systems in the Drosophila head. Neuron. 2001;32(4):657–71.

Correa E, Goodacre R. A genetic algorithm-Bayesian network approach for the analysis of metabolomics and spectroscopic data: Application to the rapid identification of Bacillus spores and classification of Bacillus species. BMC Informatics. 2011;12:33.

Creek DJ, Chokkathukalam A, Jankevics A, Burgess KEV, Breitling R, Barrett MP. Stable isotope- assisted metabolomics for network-wide metabolic pathway elucidation. Anal Chem. 2012; 84(20):8442–7.

Dallmann R, Viola AU, Tarokh L, Cajochen C, Brown SA. The human circadian metabolome. Proc Natl Acad Sci U S A. 2012;109(7):2625–9.

Davies SK, Ang JE, Revell VL, Holmes B, Mann A, Robertson FP, et al. Effect of sleep deprivation on the human metabolome. Proc Natl Acad Sci U S A. 2014;111(29):10761–6.

de Marian JJ. Observation botanique. Hist l'Academie R des Sci. 1729;35–6.

Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006;312(5771):212-7.

Ducker GS, Chen L, Morscher RJ, Ghergurovich JM, Esposito M, Teng X, et al. Reversal of cytosolic one-carbon flux compensates for loss of the mitochondrial folate pathway. Cell Metab. 2016;24(4):640–1.

Dunn OJ. Multiple comparisons among means. J Am Stat Assoc. 1961;56(293):52-64.

Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6(7):1060–83.

Ebrahim A, Lerman JA, Palsson BØ, Hyduke DR. COBRApy: Constraints-based reconstruction and analysis for Python. BMC Syst Biol. 2013;7:74.

Eckel-Mahan KL, Patel VR, Mateo S De, Orozco-Solis R, Ceglia NJ, Sahar S, et al. Reprogramming of the circadian clock by nutritional challenge. Cell. 2013;155(7):1464–78.

Eckel-Mahan KL, Patel VR, Mohney RP, Vignola KS, Baldi P, Sassone-Corsi P. Coordination of the transcriptome and metabolome by the circadian clock. Proc Natl Acad Sci U S A. 2012;109(14):5541–6.

Edery I. Circadian rhythms in a nutshell. Physiol Genomics. 2000;3(2):59–74.

Edgar RS, Green EW, Zhao Y, van Ooijen G, Olmedo M, Qin X, et al. Peroxiredoxins are conserved markers of circadian rhythms. Nature. 2012;485(7399):459–64.

Eliasson M, Rännar S, Madsen R, Donten MA, Marsden-Edwards E, Moritz T, et al. Strategy for optimizing LC-MS data processing in metabolomics: A design of experiments approach. Anal Chem. 2012;84(15):6869-76.

Emery P, So WV, Kaneko M, Hall J., Rosbash M. CRY, a Drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. Cell. 1998;95(5):669-79.

Engelen MP, Rutten EP, De Castro CL, Wouters EF, Schols AM, Deutz NE. Casein protein results in higher prandial and exercise induced whole body protein anabolism than whey protein in chronic obstructive pulmonary disease. Metabolism. 2012;61(9):1289–1300.

Eriksson L, Johansson E, Kettaneh-Wold N, Wikström C, Wold S. Design of Experiments, Principles and Applications, 2nd ed. Umeå: Umetrics Academy. 2006

Esquirol Y, Perret B, Ruidavets JB, Marquie JC, Dienne E, Niezborala M, et al. Shift work and cardiovascular risk factors: New knowledge from the past decade. Arch Cardiovasc Dis. 2011;104(12):636–68.

Fan J, Kamphorst JJ, Mathew R, Chung MK, White E, Shlomi T, et al. Glutamine-driven oxidative phosphorylation is a major ATP source in transformed mammalian cells in both normoxia and hypoxia. Mol Syst Biol. 2013;9:712.

Fan J, Ye J, Kamphorst JJ, Shlomi T, Thompson CB, Rabinowitz JD. Quantitative flux analysis reveals folate-dependent NADPH production. Nature. 2014;510(7504):298–302.

Fan TW, Lane AN, Higashi RM, Yan J. Stable isotope resolved metabolomics of lung cancer in a SCID mouse model. Metabolomics. 2011;7(2):257–69.

Feist AM, Herrgrard MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microbial organisms. Nat Rev Microbiol. 2009;7(2):129–43.

Feng D, Liu T, Sun Z, Bugge A, Mullican SE, Alenghat T, et al. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. Science. 2011;331(6022):1315–9.

Fernandez R, Tabarini D, Azpiazu N, Frasch M, Schlessinger J. The Drosophila insulin receptor homolog: A gene essential for embryonic development encodes two receptor isoforms with different signaling potential. EMBO J. 1995;14(14):3373–84.

Fischer E, Sauer U. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry Escherichia coli. J Biol Chem. 2003;278(47):46446–51.

Fouquet R, Pearson PJG. Seven centuries of energy services: The price and use of light in the United Kingdom (1300-2000). Energy J. 2006;27(1):139–77.

Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, et al. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. Nature. 2011;477(7363):225–8.

Garofalo RS. Genetic analysis of insulin signaling in Drosophila. Trends Endocrinol Metab. 2002;13(3):156–62.

Giavalisco P, Hummel J, Lisec J, Inostroza AC, Catchpole G, Willmitzer L. High-resolution direct infusion-based mass spectrometry in combination with whole 13C metabolome isotope labeling allows unambiguous assignment of chemical sum formulas. Anal Chem. 2008; 80(24):9417–25.

Giebultowicz JM. Molecular mechanism and cellular distribution of insect circadian clocks. Annu Rev Entomol. 2000;45:769–93.

Gika HG, Wilson ID, Theodoridis GA. LC-MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives. J Chromatogr B Analyt Technol Biomed Life Sci. 2014;966:1-6.

Gika HG, Theodoridis GA, Vrhovsek U, Mattivi F. Quantitative profiling of polar primary metabolites using hydrophilic interaction ultrahigh performance liquid chromatography-tandem mass spectrometry. J Chromatogr A. 2012;1259:121–7.

Gill S, Le HD, Melkani GC, Panda S. Time-restricted feeding attenuates age-related cardiac decline in Drosophila. Science. 2015;347(6227):1265–9.

Giskeødegård GF, Davies SK, Revell VL, Keun H, Skene DJ. Diurnal rhythms in the human urine metabolome during sleep and total sleep deprivation. Sci Rep. 2015;5:14843.

Glossop NR, Hardin PE. Central and peripheral circadian oscillator mechanisms in flies and mammals. J Cell Sci. 2002;115(17):3369–77.

Gogna N, Singh VJ, Sheeba V, Dorai K. NMR-based investigation of the Drosophila melanogaster metabolome under the influence of daily cycles of light and temperature. Mol Biosyst. 2015;11(12):3305–15.

Goldberg RN, Tewari YB, Bhat TN. Thermodynamics of enzyme-catalyzed reactions - A database for quantitative biochemistry. Bioinformatics. 2004;20(16):2874–7.

Greenough WB, Crespin SR, Steinberg D. Infusion of long-chain fatty acid anions by continuous-flow centrifugation. J Clin Invest. 1969;48(1):1923–33.

Géminard C, Rulifson EJ, Léopold P. Remote control of insulin secretion by fat cells in Drosophila. Cell Metab. 2009;10(3):199–207.

Hackett SR, Zanotelli VR, Xu W, Goya J, Park JO, Perlman DH, et al. Systems-level analysis of mechanisms regulating yeast metabolic flux. Science. 2016;354(6311):aaf2786.

Hao Z, Xiao B, Weng NJ. Impact of column temperature and mobile phase components on selectivity of hydrophilic interaction chromatography (HILIC). J Sep Sci. 2008;31(9):1449-64.

Hardin PE. Molecular genetic analysis of circadian timekeeping in Drosophila. Adv Genet. 2011;74:141-73.

Hartl WH, Miyoshi H, Jahoor F, Klein S, Elahi D, Wolfe RR. Bradykinin attenuates glucagon-induced leucine oxidation in humans. Am J Physiol. 1990; 259(2):E239–E245.

Hatori M, Vollmers C, Zarrinpar A, Ditacchio L, Eric A, Gill S, et al. Time restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high fat diet. Cell Metab. 2012;15(6):848–60.

Helfrich-Förster C. Robust circadian rhythmicity of Drosophila melanogaster requires the presence of lateral neurons: a brain-behavioral study of disconnected mutants. J Comp Physiol. A. 1998;182(4):435–53.

Helfrich-Förster C. Differential control of morning and evening components in the activity rhythm of Drosophila melanogaster--sex-specific differences suggest a different quality of activity. J Biol Rhythms. 2000;15(2):135-54.

Hendricks JC, Finn SM, Panckeri KA, Chavkin J, Williams JA, Sehgal A, et al. Rest in Drosophila is a sleep-like state. Neuron. 2000;25(1):129–38.

Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. Biophys J. 2007;92(5):1792–805.

Hildebrandt A, Bickmeyer I, Kuhnlein RP. Reliable Drosophila body fat quantification by a coupled colorimetric assay. PLoS One. 2011;6(9):e23796.

Hoffman JM, Soltow QA, Li S, Sidik A, Jones DP, Promislow DE. Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, Drosophila melanogaster. Aging Cell. 2014;13(4):596–604.

Horning EC, Horning MG. Metabolic profiles: gas-phase methods for analysis of metabolites. Clin Chem. 1971;17(8):802–9.

Huang X, Chen YJ, Cho K, Nikolskiy I, Crawford PA, Patti GJ. X(13)CMS: Global tracking of isotopic labels in untargeted metabolomics. Anal Chem. 2014 Feb 4;86(3):1632–9.

Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, et al. Harmonics of circadian gene transcription in mammals. PLoS Genet. 2009;5(4):e1000442.

Hughes ME, Hogenesch JB, Kornacker K. JTK_CYCLE: an efficient non-parametric algorithm for detecting rhythmic components in genome-scale datasets. J Biol Rhythm. 2010;25(5):372–80.

Hunter-Ensor M, Ousley A, Sehgal A. Regulation of the Drosophial protein timeless suggests a mechanism for resetting the clock by light. Cell. 1996;84(5):677-85.

Inoue S, Honda K, Komoda Y. Sleep as neuronal detoxification and restitution. Behav Brain Res. 1995;69(1–2):91–6.

Ivanisevic J, Zhu Z, Plate L, Tautenhahn R, Chen S, O'Brien PJ, et al. Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism. Anal Chem. 2013;85(14):6876-84.

Iwanaga H, Yano M, Miki H, Okada K, Azama T, Takiguchi S, et al. Per2 gene expressions in the suprachiasmatic nucleus and liver differentially respond to nutrition factors in rats. JPEN J Parenter Enteral Nutr. 2005;29(3):157–61.

Jakubowicz D, Barnea M, Wainstein J, Froy O. High caloric intake at breakfast vs. dinner differentially influences weight loss of overweight and obese women. Obesity (Silver Spring). 2013;21(12):2504–12.

Jay SM, Dawson D, Lamond N. Train drivers' sleep quality and quantity during extended relay operations. Chronobiol Int. 2006;23(6):1241–52.

Kaasik K, Kivimäe S, Allen JJ, Chalkley RJ, Huang Y, Baer K, et al. Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. Cell Metab. 2013;17(2):291–302.

Kalsbeek A, van der Spek R, Lei J, Endert E, Buijs RM, Fliers E. Circadian rhythms in the hypothalamo-pituitary-adrenal (HPA) axis. Mol Cell Endocrinol. 2012;349(1):20–9.

Kasukawa T, Sugimoto M, Hida A, Minami Y, Mori M, Honma S. Human blood metabolite timetable indicates internal body time. Proc Natl Acad Sci U S A. 2012;109(37):15036–41.

Katewa SD, Akagi K, Bose N, Rakshit K, Camarella T, Zheng X, et al. Peripheral circadian clocks mediate dietary restriction-dependent changes in lifespan and fat metabolism in Drosophila. Cell Metab. 2016;23(1):143–54.

Kayser MS, Yue Z, Sehgal A. A critical period of sleep for development of courtship circuitry and behavior in Drosophila. Science. 2014;344(6181):269–74.

Keegan KP, Pradhan S, Wang JP, Allada R. Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes. PLoS Comput Biol. 2007;3(11):e208.

Keene AC, Duboue ER, McDonald DM, Dus M, Suh GS, Waddell S, Blau J. Clock and cycle limit starvation-induced sleep loss in drosophila. Curr Biol. 2010; 20(13):1209–15.

Kida K, Nishio T, Yokozawa T, Nagai K, Nakagawa H. The circadian change of gluconeogenesis in the liver in vivo in fed rats. J Biochem. 1980;88(4):1009–13.

Kim EY, Jeong EH, Park S, Jeong H, Edery I. A role for O-GlcNAcylation in setting circadian clock speed. Genes Dev. 2012;26(5):490–502.

Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem substance and compound databases. Nucleic Acids Res. 2016;44(D1):D1202–13.

Kim SK, Rulifson EJ. Conserved mechanisms of glucose sensing and regulation by Drosophila corpora cardiaca cells. Nature. 2004;431(7006):316–20.

Kivilompolo M, Öhnrberg L, Orešič M, Hyötyläinen T. Rapid quantitative analysis of carnitine and acylcarnitines by ultra-high performance-hydrophilic interaction liquid chromatography-tandem mass spectrometry. Chromatogr A. 2013;1292:189-94.

Klowden MJ. Physiological systems in insects. Amsterdam: Elsevier/AP;2013.

Kondratov RV, Kondratova AA, Gorbacheva VY, Vykhovanets OV, Antoch MP. Early aging and age-related pathologies in mice deficient in BMAL1, the core componentof the circadian clock. Genes Dev. 2006;20(14):1868–73.

Konopka RJ, Benzer S. Clock mutants of Drosophila melanogaster. Proc Natl Acad Sci U S A. 1971;68(9):2112–6.

Kostic N, Dotsikas Y, Malenovic A, Stojanovic JB, Rakic T, Ivanovic D, Medenica MJ. Stepwise optimization approach for improving LC-MS/MS analysis of zwitterionic antiepileptic drugs with implementation of experimental design. Mass Spectrom. 2013;48(7):875-4.

Kotlowska A. Application of Chemometric Techniques in Search of Clinically Applicable Biomarkers of Disease. Drug Dev Res. 2014;75(5):283–90.

Krishnan N, Davis AJ, Giebultowicz JM. Circadian regulation of response to oxidative stress in Drosophila melanogaster. Biochem Biophys Res Commun. 2008;374(2):299–303.

Krishnan N, Kretzschmar D, Rakshit K, Chow E, Giebultowicz. The circadian clock gene period extends healthspan in aging Drosophila melanogaster. Aging (Albany NY). 2009;1(11):937–48.

Krishnan N, Rakshit K, Chow ES, Wentzell JS, Kretzschmar D, Giebultowicz JM. Loss of circadian clock accelerates aging in neurodegeneration-prone mutants. Neurobiol Dis. 2012;45(3):1129-35.

Krupp JJ, Kent C, Billeter JC, Azanchi R, So AK, Schonfeld JA, et al. Social experience modifies pheromone expression and mating behavior in male Drosophila melanogaster. Curr Biol. 2008;18(18):1373–83.

Kumar D, Wingate D, Ruckebusch Y. Circadian variation in the propagation velocity of the migrating motor complex. Gastroenterology. 1986;91(4):926–30.

Kume K, Kume S, Park SK, Hirsh J, Jackson FR. Dopamine is a regulator of arousal in the fruit fly. J Neurosci. 2005;25(32):7377–84.

Köhler E, Barrach HJ, Neubert D. Inhibition of NADP dependent oxidoreductases by the 6-aminonicotinamide analogue of NADP. FEBS Lett. 1970;6(3):225–8.

La Fleur SE, Kalsbeek A, Wortel J, Buijs RM. A suprachiasmatic nucleus generated rhythm in basal glucose concentrations. J Neuroendocrinol. 1999;11(8):643–52.

La Fleur SE, Kalsbeek A, Wortel J, Fekkes ML, Buijs RM. A daily rhythm in glucose tolerance. Diabetes. 2001;50(6):1237–43.

Lamia KA, Sachdeva UM, Ditacchio L, Williams EC, Alvarez JG, Egan DF, et al. AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. Science. 2009;326(5951):437–40.

Laposky A, Easton A, Dugovic C, Walisser J, Bradfield C, Turek F. Deletion of the mammalian circadian clock gene BMAL1/Mop3 alters baseline sleep architecture and the response to sleep deprivation. Sleep. 2005;28(4):395–409.

Laye MJ, Tran V, Jones DP, Kapahi P, Promislow DEL. The effects of age and dietary restriction on the tissue-specific metabolome of Drosophila. Aging Cell. 2015;14(5):797–808.

Lee G, Park JH. Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in Drosophila melanogaster. Genetics. 2004;167(1):311–23.

Lemmer B. Discoveries of rhythms in human biological functions: a historical review. Chronobiol Int. 2009;26(6):1019–68.

Leopold P, Perrimon N. Drosophila and the genetics of the internal milieu. Nature. 2007;450(7167):186–8.

Lewis CA, Parker SJ, Fiske BP, McCloskey D, Gui DY, Green CR, et al. Tracing Compartmentalized NADPH Metabolism in the Cytosol and Mitochondria of Mammalian Cells. Mol Cell. 2014;55(2):253–63.

Li M, Ruan H, Hughes ME, Lee J, Singh JP, Jones SP, et al. O-GlcNAc signaling entrains the circadian clock by inhibiting BMAL1/CLOCK ubiquitination. Cell Metab. 2013;17(2):303–10.

Lin Y, Han M, Shimada B, Wang L, Gibler TM, Amarakone A, et al. Influence of the period-dependent circadian clock on diurnal, circadian, and aperiodic gene expression in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2002;99(14):9562–7.

Liu S, Brown JD, Stanya KJ, Homan E, Leidl M, Inouye K, et al. A diurnal serum lipid integrates hepatic lipogenesis and peripheral fatty acid use. Nature. 2013;502(7472):550–4.

Livingstone, MS, Tempel BL. Genetic dissection of mono- amine neurotransmitter synthesis in Drosophila. Nature. 1983;303(5912):67–70.

Li Y, Li G, Görling B, Luy B, Du J, Yan J. Integrative analysis of circadian transcriptome and metabolic network reveals the role of de novo purine synthesis in circadian control of cell cycle. PLoS Comput Biol. 2015;11(2):e1004086.

Loeb J, Northrop JH. On the influence of food and temperature on the duration of life. J Biol Chem. 1917;32:103–21.

Longo VD, Panda S. Fasting, circadian rhythms, and time-restricted feeding in healthy lifespan. Cell Metab. 2016;23(6):1048–59.

Lowrey PL, Takahashi JS. Genetics of circadian rhythms in mammalian model organisms. Adv Genet. 2011;74:175–230.

Luckhaupt SE, Tak S, Calvert GM. The prevalence of short sleep duration by industry and occupation in the National Health Interview Survey. Sleep. 2010;33(2):149–59.

Lv H, Palacios G, Hartil K, Kurland IJ. Advantages of tandem LC-MS for the rapid assessment of tissue-specific metabolic complexity using a pentafluorophenylpropyl stationary phase. J Proteome Res. 2011;10(4):2104-12.

Maguire SE, Rhoades S, Chen WF, Sengupta A, Yue Z, Lim JC, et al. Independent effects of g-aminobutyric acid transaminase (GABAT) on metabolic and sleep homeostasis. J Biol Chem. 2015;290(33):20407-16.

Mapstone M, Cheema AK, Fiandaca MS, Zhong X, Mhyre TR, MacArthur LH, et al. Plasma phospholipids identify antecedent memory impairment in older adults. Nat Med. 2014;20(4):415-8.

Marcheva B, Ramsey KM, Buhr ED, Kobayashi Y, Su H, Ko CH, et al. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinemia and diabetes. Nature. 2010;466(7306):627–31.

Masri S, Papagiannakopoulos T, Kinouchi K, Liu Y, Cervantes M, Baldi P, et al. Lung adenocarcinoma distally rewires hepatic circadian homeostasis. Cell. 2016;165(4):896–909.

Masri S, Patel VR, Eckel-Mahan KL, Peleg S, Forne I, Ladurner AG, et al. Circadian acetylome reveals regulation of mitochondrial metabolic pathways. Proc Natl Acad Sci U S A. 2013;110(9):3339–44.

Mattaini KR, Sullivan MR, Vander Heiden MG. The importance of serine metabolism in cancer. J Cell Biol. 2016;214(3):249–57.

Mauvoisin D, Wang J, Jouffe C, Martin E, Atger F, Waridel P, et al. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. Proc Natl Acad Sci U S A. 2014;111(1):167–72.

Mavrovouniotis ML. Group contributions for estimating standard gibbs energies of formation of biochemical compounds in aqueous solution. Biotechnol Bioeng. 1990;36(10):1070-82.

Maywood ES, Chesham JE, O'Brien J a, Hastings MH. A diversity of paracrine signals sustains molecular circadian cycling in suprachiasmatic nucleus circuits. Proc Natl Acad Sci U S A. 2011;108(34):14306–11.

McDonald MJ, Rosbash M. Microarray analysis and organization of circadian gene expression in Drosophila. Cell. 2001;107(5):567–78.

McKenna MC, Sonnewald U. GABA alters the metabolic fate of [U-13C]glutamate in cultured cortical astrocytes. J Neurosci Res. 2005;79(1–2):81–7.

Minami Y, Kasukawa T, Kakazu Y, Iigo M, Sugimoto M, Ikeda S, et al. Measurement of internal body time by blood metabolomics. Proc Natl Acad Sci U S A. 2009;106(24):9890–5.

Moore RY, Eichler VB. Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. Brain Res. 1972;42(1):201–6.

Moore RY. The suprachiasmatic nucleus and the organization of a circadian system. Trends Neurosci. 1982;5:404–7.

Moseley HN. Correcting for the effects of natural abundance in stable isotope resolved metabolomics experiments involving ultra-high resolution mass spectrometry. BMC Bioinformatics. 2010;11:139.

Munger J, Bennett BD, Parikh A, Feng XJ, McArdle J, Rabitz HA, et al. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nat Biotechnol. 2008;26(10):1179–86.

Musselman LP, Fink JL, Narzinski K, Ramachandran PV, Hathiramani SS, Cagan RL, Baranski TJ. A high-sugar diet produces obesity and insulin resistance in wild-type Drosophila. Dis Model Mech. 2011;4(6):842–9.

Musselman LP, Fink JL, Ramachandran PV, Patterson BW, Okunade AL, Maier E, et al. Role of fat body lipogenesis in protection against the effects of caloric overload in Drosophila. J Biol Chem. 2013;288(12):8028–42.

Myers, MP, Wager-Smith, K, Rothenfluh-Hilfiker, A, Young MW. Light-Induced degradation of TIMELESS and entrainment of the Drosophila circadian clock. Science. 1996;271(5256):1736-40.

Myers EM, Yu J, Sehgal A. Circadian control of eclosion: interaction between a central and peripheral clock in Drosophila melanogaster. Curr Biol 2003;13(6):526–33.

Nall A, Sehgal A. Monoamines and sleep in Drosophila. Behav Neurosci. 2014;128(3):264–72.

Napoles MO, Steenbergen RM. Analysis of axle and vehicle load properties through Bayesian Networks based on Weigh-in-Motion data. Reliab Eng Syst Safe. 2014;125:153-64.

Naylor E, Bergmann BM, Krauski K, Zee PC, Takahashi JS, Vitaterna MH, et al. The circadian clock mutation alters sleep homeostasis in the mouse. J Neurosci. 2000;20(21):8138–43.

New L, Chan EC. Evaluation of BEH C18, BEH HILIC, and HSS T3 (C18) column chemistries for the UPLC-MS-MS analysis of glutathione, glutathione disulfide, and ophthalmic acid in mouse liver and human plasma. J Chromatogr Sci. 2008;46(3):209-14.

Nguyen HP, Schug KA. The advantages of ESI-MS detection in conjunction with HILIC mode separations: Fundamentals and applications. J Sep Sci. 2008;31(9):1465-80.

Nicholson, SD. An Introduction to Metabolic Pathways. Oxford and Edinburgh: Blackwell Scientific Publications. 1970.

Nicholson JK, Lindon JC. Systems biology: Metabonomics. Nature. 2008;455(7216):1054-6.

Oosterman JE, Kalsbeek A, la Fleur SE, Belsham DD. Impact of nutrients on circadian rhythmicity. Am J Physiol Regul Integr Comp Physiol. 2015;308(5):R337–50.

Opperhuizen AL, van Kerkhof LW, Proper KI, Rodenburg W, Kalsbeek A. Rodent models to study the metabolic effects of shiftwork in humans. Front Pharmacol. 2015;6:50.

Orr WC, Radyuk SN, Sohal RS. Involvement of redox state in the aging of Drosophila melanogaster. Antioxid Redox Signal. 2013;19(8):788–803.

Paglia G, Hrafnsdóttir S, Magnúsdóttir M, Fleming RM, Thorlacious S, Palsson BØ, Thiele I. Monitoring metabolites consumption and secretion in cultured cells using ultra-performance liquid chromatography quadrupole-time of flight mass spectrometry (UPLC-Q-ToF-MS). Anal Bioanal Chem. 2012;402(3):1183-98.

Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. Cell. 2002 May 3;109(3):307–20.

Pan X, Hussain MM. Clock is important for food and circadian regulation of macronutrient absorption in mice. J Lipid Res. 2009;50(9):1800–13.

Pan X, Hussain MM. Diurnal regulation of microsomal triglyceride transfer protein and plasma lipid levels. J Biol Chem. 2007;282(34):24707–19.

Papagiannakopoulos T, Bauer MR, Davidson SM, Heimann M, Subbaraj L, Bhutkar A, et al. Circadian rhythm disruption promotes lung tumorigenesis. Cell Metab. 2016;24(2):324-31.

Park J, Lee SB, Lee S, Kim Y, Song S, Kim S, et al. Mitochondrial dysfunction in Drosophila PINK1 mutants is complemented by parkin. Nature. 2006;441(7097):1157–61.

Park JO, Rubin SA, Xu YF, Amador-Noguez D, Fan J, Shlomi T, et al. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat Chem Biol. 2016;12(7):482–9.

Paschos GK, Ibrahim S, Song WL, Kunieda T, Grant G, Reyes TM, et al. Obesity in mice with adipocyte-specific deletion of clock component Arntl. Nat Med. 2012;18(12):1768–77.

Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proc Natl Acad Sci U S A. 1971;68(10):2374-6.

Peek CB, Affinati AH, Ramsey KM, Kuo H-Y, Yu W, Sena L a, et al. Circadian clock NAD+ cycle drives mitochondrial oxidative metabolism in mice. Science. 2013;342(6158):1243417.

Perelis M, Marcheva B, Ramsey KM, Schipma MJ, Hutchison AL, Taguchi A, et al. Pancreatic ß-cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. Science. 2015;350(6261):aac4250.

Peschke E, Peschke D. Evidence for a circadian rhythm of insulin release from perifused rat pancreatic islets. Diabetologia. 1998;41(9):1085–92.

Pitsouli C, Perrimon N. Developmental biology: Our fly cousins' gut. Nature. 2008;454(7204):592–3.

Pittendrigh CS. Circadian rhythms and the circadian organization of living systems. Cold Spring Harb Symp Quant Biol. 1960;25:159–84.

Pizer LI. The pathway and control of serine biosynthesis in Escherichia coli. J Biol Chem. 1963;238(12):3934-44.

Plautz JD, Kaneko M, Hall JC, Kay SA. Independent photoreceptive circadian clocks throughout Drosophila. Science. 1997;278(5343):1632–5.

Postic C, Dentin R, Denechaud PD, Girard J. ChREBP, a transcriptional regulator of glucose and lipid metabolism. Annu Rev Nutr. 2007;27:179–92.

Puttonen S, Harma M, Hublin C. Shift work and cardiovascular disease - Pathways from circadian stress to morbidity. Scand J Work Environ Heal. 2010;36(2):96–108.

Qu H, Konradsen JR, van Hengel M, Wolt S, Sonnewald U. Effect of glutamine and GABA on [U-(13)C]glutamate metabolism in cerebellar astrocytes and granule neurons. J Neurosci Res. 2001;66(5):885–90.

Ramautar R, Somsen G, de Jong G. CE-MS in metabolomics. Electrophoresis. 2009;30(1):276–91.

Ramsey KM, Yoshino J, Brace CS, Abrassart D, Kobayashi Y, Marcheva B, et al. Circadian clock feedback cycle through NAMPT-Mediated NAD+ biosynthesis. Science. 2009;324(5927):651–4.

Rayne RC, O'Shea M. Reconstitution of adipokinetic hormone biosynthesis in vitro indicates steps in prohormone processing. Eur J Biochem. 1994;219(3):781–9.

Reddy AB, Karp NA, Maywood ES, Sage EA, Deery M, O'Neill JS, et al. Circadian orchestration of the hepatic proteome. Curr Biol. 2006;16(11):1107–15.

Regestein QR, Monk TH. Delayed sleep phase syndrome: A review of its clinical aspects. Am J Psychiatry. 1995;152(4):602–8.

Reilly DF, Westgate EJ, FitzGerald GA. Peripheral circadian clocks in the vasculature. Arterioscler Thromb Vasc Biol. 2007;27(8):1694–705.

Rey G, Valekunja UK, Feeney KA, Wulund L, Milev NB, Stangherlin A, et al. The pentose phosphate pathway regulates the circadian clock. Cell Metab. 2016;24(3):462-73.

Rhoades SD, Sengupta A, Weljie AM. Time is ripe: maturation of metabolomics in chronobiology. Curr Opin Biotechnol. 2017;43:70–6.

Rhoades SD, Weljie AM. Comprehensive optimization of LC–MS metabolomics methods using design of experiments (COLMeD). Metabolomics. 2016;12(12):183.

Riter LS, Vitek O, Gooding KM, Hodge BD, Julian RK. Statistical design of experiments as a tool in mass spectrometry. J Mass Spectrom. 2005;40(5):565-79.

Saltiel AR, Kahn CR. Insulin signaling and the regulation of glucose and lipid metabolism. Nature. 2001;414(6865):799–806.

Sampsonidis I, Witting M, Koch W, Virgillou C, Gika HG, Schmitt-Kopplin P, Theodoridis GA. Computational analysis and ratiometric comparison approaches aimed to assist column

selection in hydrophilic interaction liquid chromatography-tandem mass spectrometry targeted metabolomics. J Chromatogr A. 2015;1406:145-55.

Sauer U. Metabolic networks in motion: 13C-based flux analysis. Mol Syst Biol. 2006;2:62.

Scheer FA, Hilton MF, Mantzoros CS, Shea SA. Adverse metabolic and cardiovascular consequences of circadian misalignment. Proc Natl Acad Sci U S A. 2009;106(11):4453–8.

Sengupta A, Krishnaiah S, Rhoades S, Growe J, Slaff B, Venkataraman A, et al. Deciphering the duality of clock and growth metabolism in a cell autonomous system using NMR profiling of the secretome. Metabolites. 2016;6(3):E23.

Shang Y, Haynes P, Pírez N, Harrington KI, Guo F, Pollack J, et al. Imaging analysis of clock neurons reveals light buffers the wake-promoting effect of dopamine. Nat Neurosci. 2011;14(7):889–95.

Shi G. Application of co-eluting structural analog internal standards for expanded linear dynamic range in liquid chromatography/electrospray mass spectrometry. Rapid Commun Mass Spectrom. 2003;17(3):202–6.

Shin SC, Kim SH, You H, Kim B, Kim AC, Lee KA, et al. Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science. 2011;334(6056):670-4.

Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. Nat Biotechnol. 2008;26(9):1003–10.

Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. Ther Drug Monit. 2005;27(6):747-51.

Spencer RL, Kim PJ, Kalman BA, Cole MA. Evidence for mineralocorticoid receptor facilitation of glucocorticoid receptor-dependent regulation of hypothalamic-pituitary-adrenal axis activity. Endocrinology. 1998;139(6):2718–26.

Staubli F, Jorgensen TJ, Cazzamali G, Williamson M, Lenz C, Sondergaard L, et al. Molecular identification of the insect adipokinetic hormone receptors. Proc Natl Acad Sci U S A. 2002;99(6):3446–51.

Steele R. Influences of glucose loading and of injected insulin on hepatic glucose output. Ann N Y Acad Sci. 1959;82:420–430.

Suh S-H, Casazza GA, Horning MA, Miller BF, Brooks GA. Effects of oral contraceptives on glucose flux and substrate oxidation rates during rest and exercise. J Appl Physiol. 2003;94(1):285–94.

Sutin E, Kilduff T. Circadian and light-induced expression of immediate early gene mRNAs in the rat suprachiasmatic nucleus. Brain Res Mol Brain Res. 1992;15(3–4):281–90.

Székely GY, Henriques B, Gil M, Ramos A, Alvarez CJ. Design of experiments as a tool for LC-MS/MS method development for the trace analysis of the potentially genotoxic 4-dimethylaminopyridine impurity in glucocorticoids. Pharm Biomed Anal. 2012;70:251-8.

Taghert PH, Shafer OT. Mechanisms of clock output in the Drosophila circadian pacemaker system. J Biol Rhythms. 2006;21(6):445–57.

Teclemariam-Mesbah R, Ter Horst G, Postema F, Wortel J, Buijs R. Anatomical demonstration of the suprachiasmatic nucleus–pineal pathway. J Comp Neurol. 1999;406(2):171–82.

Teleman AA, Chen YW, Cohen SM. Drosophila Melted modulates FOXO and TOR activity. Dev Cell. 2005;9(2):271–81.

Tennessen JM, Barry WE, Cox J, Thummel CS. Methods for studying metabolism in Drosophila. Methods. 2014;68(1):105–15.

Tennessen JM, Bertagnolli NM, Evans J, Sieber MH, Cox J, Thummel CS. Coordinated metabolic transitions during Drosophila embryogenesis and the onset of aerobic glycolysis. G3 (Bethesda). 2014;4(5):839–50.

Tepper N, Noor E, Amador-Noguez D, Haraldsdottir HS, Milo R, Rabinowitz J, et al. Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. PLoS One. 2013;8(9):e75370.

Theodoridis GA, Gika HG, Want EJ, Wilson ID. Liquid chromatography-mass spectrometry based global metabolite profiling: A review. Anal Chim Acta. 2012;711:7–16.

Thurley K, Herbst C, Wesener F, Koller B, Wallach T, Maier B, et al. Principles for circadian orchestration of metabolic pathways. Proc Nat Acad Sci U S A. 2017;114(7):1572–7.

Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. J Proteome Res. 2007;6(2):469-79.

Tsai JY, Kienesberger PC, Pulinilkunnil T, Sailors MH, Durgan DJ, Villegas-Montoya C, et al. Direct regulation of myocardial triglyceride metabolism by the cardiomyocyte circadian clock. J Biol Chem. 2010;285(5):2918–29.

Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, Laposky A, et al. Obesity and metabolic syndrome in circadian clock mutant mice. Science. 2005;308(5724):1043–5.

Van Alphen B, Yap MH, Kirszenblat L, et al. A dynamic deep sleep stage in Drosophila. J Neurosci. 2013; 33(16):6917–27.

Van Den Berg R, Mook-Kanamori DO, Donga E, Van Dijk M, Van Dijk JG, Lammers GJ, et al. A single night of sleep curtailment increases plasma acylcarnitines: Novel insights in the relationship between sleep and insulin resistance. Arch Biochem Biophys. 2016;589:145–51.

van der Greef J, Smilde AK. Symbiosis of chemometrics and metabolomics: past, present, and future. J Chemom. 2005;19(5–7):376–86.

Van Someren EJ, Lijzenga C, Mirmiran M, Swaab DF. Long-term fitness training improves the circadian rest-activity rhythm in healthy elderly males. J Biol Rhythms. 1997;12(2):146–56.

Vollmers C, Gill S, DiTacchio L, Pulivarthy SR, Le HD, Panda S. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc Natl Acad Sci U S A. 2009;106(50):21453–8.

Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc. 2010;5(6):1005-18.

Weisz, PB. Diffusion and chemical transformation. Science. 1973;179(4072):433-40.

Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, Wiechert W. 13CFLUX2 - High-performance software suite for 13C-metabolic flux analysis. Bioinformatics. 2013;29(1):143–5.

Weljie AM, Meerlo P, Goel N, Sengupta A, Kayser MS, Abel T, et al. Oxalic acid and diacylglycerol 36:3 are cross-species markers of sleep debt. Proc Natl Acad Sci U S A. 2015;112(8):2569–74.

Wiechert W, Nöh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. Curr Opin Biotechnol. 2013;24(6):979–86.

Wijnen H, Naef F, Boothroyd C, Claridge-Chang A, Young MW. Control of daily transcript oscillations in Drosophila by light and the circadian clock. PLoS Genet. 2006;2(3):e39.

Williams RJ. Individual metabolic patterns and human disease: An exploratory study utilizing predominantly paper chromatographic methods. The University of Texas Publication. 1951.

Winder CL, Dunn WB, Goodacre R. TARDIS-based microbial metabolomics: Time and relative differences in systems. Trends Microbiol. 2011;19(7):315–22.

Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41:D801-7.

Wisor JP, O'Hara BF, Terao A, Selby CP, Kilduff TS, Sancar A, et al. A role for cryptochromes in sleep regulation. BMC Neurosci. 2002;3:20.

Wolfe RR, Chinkes DL. Isotope Tracers in Metabolic Research. 2nd ed. John Wiley & Sons, Inc.: Hoboken, New Jersey, USA, 2005.

Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB. MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics. 2016;32(21):3351–3.

Wyse C, Biello SM, Gill JMR. The bright-nights and dim-days of the urban photoperiod: Implications for circadian rhythmicity, metabolism and obesity. Ann Med. 2014;46(5):253–63.

Xu K, Diangelo JR, Hughes ME, Hogenesch JB, Sehgal A. The circadian clock interacts with metabolic physiology to influence reproductive fitness. Cell Metab. 2011;13(6):639–54.

Xu K, Zheng X, Sehgal A. Regulation of feeding and metabolism by neuronal and peripheral clocks in Drosophila. Cell Metab. 2008;8(4):289–300.

Yamada T, Letunic I, Okuka S, Kanehisa M, Bork P. iPath2.0: interactive pathway explorer. 2011;39:W412-5.

Yamamoto A, Zwarts L, Callaerts P, Norga K, Mackay TF, Anholt RR. Neurogenetic networks for startle-induced locomotion in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2008;105(34):12393–8.

Yamazaki S, Ishida Y, Inouye SIT. Circadian rhythms of adenosine triphosphate contents in the suprachiasmatic nucleus, anterior hypothalamic area and caudate putamen of the rat - negative correlation with electrical activity. Brain Res. 1994;664(1–2):237–40.

Yang G, Paschos G, Curtis AN, Musiek ES, McLoughlin SC, FitzGerald GA. Knitting up the raveled sleave of care. Sci Transl Med. 2013;5(212);212rv3.

Yang R, Su Z. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics. 2010;26(12):168–74.

Yizhak K, Gaude E, Le Devedec S, Waldman YY, Stein GY, van de Water B, et al. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. eLife. 2014;3:e03641.

Young M, McGinnis G. Circadian regulation of metabolic homeostasis: causes and consequences. Nat Sci Sleep. 2016;8:163-80.

Yuan M, Breitkopf SB, Yang X, Asara JM. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. Nat Protoc. 2012;7(5):872-81.

Yuan Q, Joiner WJ, Sehgal A. A sleep-promoting role for the Drosophila serotonin receptor 1A. Curr Biol. 2006;16(11):1051–62.

Zamboni N, Fendt SM, Rühl M, Sauer U. 13C-based metabolic flux analysis. Nat Protoc. 2009; 4(6):878–92.

Zamboni N, Fischer E, Sauer U. FiatFlux--a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics. 2005;6:209.

Zamboni N, Saghatelian A, Patti GJ. Defining the metabolome: Size, flux, and regulation. Mol Cell. 2015;58(4):699–706.

Zhang H, Liu J, Li CR, Momen B, Kohanski RA, Pick L. Deletion of Drosophila insulin-like peptides causes growth defects and metabolic abnormalities. Proc Natl Acad Sci U S A. 2009;106(46):19617–22.

Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. Proc Natl Acad Sci U S A. 2014;111(45):16219-24.

Zheng H, Clausen MR, Dalsgaard KT, Mortensen G, Bertram CH. Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic approaches. Anal Chem. 2013;85(15):7109-16.

Zhou G, Pang H, Tang Y, Yao X, Mo X, Zhu S, et al. Hydrophilic interaction ultra-performance liquid chromatography coupled with triple-quadrupole tandem mass spectrometry for highly rapid and sensitive analysis of underivatized amino acids in functional foods. Amino Acids. 2013;44(5):1293-305.

Zhou Y, Song J, Choi FF, Wu H, Qiao C, Ding L, et al. An experimental design approach using response surface techniques to obtain optimal liquid chromatography and mass spectrometry conditions to determine the alkaloids in Meconopsi species. J Chromatogr A. 2009;1216(42):7013-23.

Zwighaft Z, Aviram R, Shalev M, Rousso-Noori L, Kraut-Cohen J, Golik M, et al. circadian clock control by polyamine levels through a mechanism that declines with age. Cell Metab. 2015;22(5):874–85.